ADVANCES IN PHYSICS

A QUARTERLY SUPPLEMENT of the PHILOSOPHICAL MAGAZINE

VOLUME 7 JULY 1958 NUMBER 27

PRICE per part £1
PRICE per annum £3 15*s*. 0*d*. post free

PRINTED AND PUBLISHED BY TAYLOR & FRANCIS LTD
RED LION COURT, FLEET ST., LONDON E.C.4

Physics in Medicine and Biology

A Taylor & Francis International Journal published in association with the Hospital Physicists' Association

Contents of April, 1958

Subscription price per volume £3 10s. post free, payable in advance

4 parts per volume—£1 per part plus postage

Printed and Published by

TAYLOR & FRANCIS, LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4

Orders originating in U.S.A. and Canada should be sent to the Academic Press Inc., 111 Fifth Avenue, New York, 3, N.Y., U.S.A.

Journal of Electronics and Control

A Philosophical Magazine Associated Journal

Editor:
J. THOMSON, M.A., D.Sc., M.I.E.E., F.Inst.P.

Consultant Editor:
Professor N. F. MOTT, F.R.S.

Editorial Board:
Professor P. AIGRAIN (France)
Professor H. B. G. CASIMIR (Holland)
J. F. COALES (U.K.)
D. W. FRY (U.K.)
Dr. W. KLEIN (Germany)
Dr. R. KOMPFNER (U.S.A.)

Contents of July, 1958

Price per part £1 5s. plus postage

Price per volume £7 post free, payable in advance

6 monthly issues per volume

Printed and Published by

TAYLOR & FRANCIS LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4

Orders originating in U.S.A. and Canada should be sent to the Academic Press Inc., 111 Fifth Avenue, New York, 3, N.Y. U.S.A.

The Philosophical Magazine

First Published in 1798

Contents of August, 1958

Price per part £1 5s. plus postage

Price per annum £13 10s. post free, payable in advance

Printed and Published by

TAYLOR & FRANCIS LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4

Contents of July, 1958

Price per part £1 5s. 0d. plus postage

Subscription per volume (4 issues) £4 15s. 0d. post free, payable in advance

Printed and Published by

TAYLOR & FRANCIS LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4

Orders originating in U.S.A. and Canada should be sent to the Academic Press Inc., 111 Fifth Avenue, New York, 3, N.Y., U.S.A.

A HISTORY OF MATHEMATICS

from antiquity to the early nineteenth century

by J. F. SCOTT, B.A., D.Sc., Ph.D.
Vice-Principal of St. Mary's College, Strawberry Hill, Twickenham, Middlesex

Author of *The Scientific Work of René Descartes* (1596–1650), *Mathematical Work of John Wallis, D.D., F.R.S.* (1616–1703), and other works.

CONTENTS : Mathematics in Antiquity—Greek Mathematics—The Invention of Trigonometry—Decline of Alexandrian Science and the Revival in Europe—Mathematics in the Orient—Progress of Mathematics during the Renaissance—New Methods in Geometry—The Rise of Mechanics—the Invention of Decimal Fractions and of Logarithms—Newton and the Calculus—Taylor and Maclaurin, the Bernoullis and Euler, Related Advances—The Calculus of Variations, Probability, Projective Geometry, Non-Euclidean Geometry—Theory of Numbers—Lagrange, Legendre, Laplace, Gauss.

This volume is intended primarily to help students who desire to have a knowledge of the development of the subject but who have too little leisure to consult original works and documents. The author has availed himself of the facilities afforded by the Royal Society and other learned Societies to reproduce extracts from manuscripts and many scarce works.

9¾ × 6¾ Price 3 guineas

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

THE SCIENTIFIC WORK OF RENÉ DESCARTES 1596-1650

BY

J. F. SCOTT, B.A., M.SC., PH.D.

With a Foreword by

H. W. TURNBULL, M.A., F.R.S.

This book puts the chief mathematical and physical discoveries of Descartes in an accessible form and fills an outstanding gap upon the shelf devoted to the history of philosophy and science. There is to be found in this volume the considerable contribution that Descartes made to the physical sciences, which involved much accurate work in geometrical optics and its bearing upon the practical problem of fashioning lenses, as also the deeper problems of light and sight and colour. The careful treatment that Dr. Scott has accorded to the work of Descartes is very welcome. The book is well worth reading and will be an asset to all libraries. This publication is recommended and approved by the Publication Fund Committee of the University of London.

212 *pp.* 7 × 10 *amply illustrated*

PRICE £1–0–0 NET

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

CONTENTS

ADVANCES IN PHYSICS

A QUARTERLY SUPPLEMENT
of the
PHILOSOPHICAL MAGAZINE

VOLUME 7 | JULY 1958 | NUMBER 27

The Electron Microscopy of Crystal Lattices

By J. W. MENTER
Tube Investments Research Laboratories, Hinxton Hall, Cambridge

CONTENTS

§ 1. Introduction

Some information about the arrangement of atoms and molecules in crystals can be obtained by a study of their optical properties. Measurement of refractive index, birefringence, optical rotation, etc., gives important clues on the disposition of the molecules with respect to the external faces of the crystal. On account of the resolution limit imposed by the wavelength of the radiation used in these methods it is not of course possible to make any observations related to the dimensions of the molecular lattice. To do this we generally resort to diffraction methods, using a radiation of wavelength comparable with that of the dimensions we wish to resolve. From calculations based on the directions and intensities of the diffracted beams of either x-rays or neutrons it is possible to reconstruct a projection of the crystal lattice showing the relative positions of the atoms in the unit cell of the lattice. This procedure can also be followed using electrons, but in this case a further possibility arises as a result of the ability of a magnetic or electrostatic field to focus an electron beam. Using electron lenses suitably combined into an electron microscope it is possible, in principle, to image a crystal lattice directly. This has two important consequences in the study of the microstructure of crystals. Firstly, it should eliminate the conceptual and computational difficulties of deducing the structure of the crystal in real space from the experimental data of reciprocal space and, secondly, departures from perfection are directly imaged. The latter is particularly important since in diffraction methods the reconstructed projection is essentially an average of all the unit cells in the volume of crystal illuminated by the incident beam of radiation so that imperfections in

the lattice will in general be smoothed out. It is true that imperfections affect the diffracted beams but generally only to a second order. Diffraction spots are distorted, the intensity of reflection is affected and spots may be displaced from their true positions. Calculations have been made of the expected diffraction effects from single line imperfections (screw dislocation, Wilson 1949, edge dislocation, Suzuki 1957), planar defects (stacking fault, Paterson 1952) and point defects (Cochran and Kartha 1956 a, b, c) and some of these have been confirmed by observations on imperfect optical gratings by Willis (1957). Unfortunately, the scattered intensity from the necessarily small volume of material required to show such individual effects is so small as to make the experimental difficulties prohibitive. The aggregate effect of such disturbances of the diffracted beams can of course be studied and the difficulties of interpretation mitigated to some extent by the use of microbeams (see, for example, review by Hirsch 1956), but detailed interpretation is often ambiguous, and computation is formidable and approximate.

Clearly then, a direct imaging method has many advantages and the technical development of the electron microscope has now reached a level where this is just becoming possible. There is no suggestion that diffraction methods will eventually be superseded. Apart from basic difficulties of resolution limit in the more direct method there are many other factors such as the form and nature of the specimen that can usefully be studied in the electron microscope, which will limit its field of applicability. The purpose of this review is to summarize the progress that has been made in this relatively new field of crystal analysis.

§ 2. Present Position of Instrument

Much of the work to be described later has been carried out near the limit of resolution of currently available electron microscopes. It is therefore relevant to consider briefly the present status of the instrument and the factors that are currently limiting the instrumental resolving power obtainable. Detailed consideration of the design and construction of the electron microscope has been covered adequately in reviews by Haine (1954) and Leisegang (1956 a). It appears that in order to make the maximum use of the resolution obtainable in a properly designed and corrected objective lens, a high electron optical magnification is required and this in turn requires a particularly efficient illuminating system in order to provide a level of illumination adequate for accurate visual focusing of the image.

2.1. *Illuminating System*

The required brightness can be obtained with a conventional hairpin tungsten cathode and a single condenser lens, but the total energy input to the specimen under these conditions is unduly high and leads to thermal instability and possible damage in the specimen so that designers have introduced a double condenser lens in the illuminating system. The

magnification required on the final screen is of the order of 100 000 times and with a photographic plate some 10 cm in diameter, the field of view is no more than about 10^{-4} cm in diameter. The conventional single condenser lens when focused on the specimen illuminates an area of the order of 50×10^{-4} cm diameter. This is unnecessarily large at high magnifications and the second condenser lens has been introduced to reduce the diameter of the focused spot of illumination to a few microns. The consequent reduction in the temperature of the specimen has unfortunately enhanced another disturbing effect, the contamination of the specimen by carbon during observation, but methods of dealing with this are under development (see § 4.4).

2.2. *Objective Lens and Resolving Power*

Elementary considerations based on the wavelength of the electrons commonly used for electron microscopes (at 80 kv, $\lambda = 0{\cdot}04$ A) suggest that detail of less than atomic dimensions should be readily resolved. This theoretical resolution is not reached with electron lenses because it has proved impossible, so far, to design large aperture systems which are correctable in the way that glass objectives are for light. Very small apertures are therefore used, of the order of 10^{-2}–10^{-3} rad, and the objective lens is designed so as to minimize the combined effects of the four most important sources of error in the image, viz. diffraction error, spherical aberration, astigmatism and chromatic aberration. The objective lens provides the initial stage of magnification in the microscope of about 30 times and it is the aberrations of this lens which largely determine the quality of the image. From the point of view of maximizing the performance of the objective lens, diffraction error and spherical aberration are of the most fundamental importance. An emitting point object is imaged by a perfect lens as an Airy diffraction pattern, the diameter of the so-called Airy disc or central maximum being $1{\cdot}22\lambda/\sin\alpha$ where 2α is the aperture angle of rays from the point to the lens in the object space. A lens with spherical aberration images a point object as a disc with minimum confusion diameter given by $\frac{1}{2}C_s\alpha^3$ where C_s is the spherical aberration constant. There is clearly an optimum value of α which minimizes the combined effect of these two sources of aberration. No completely satisfactory analysis has been made to show how these two errors should be combined. The method of combination determines the values of the constants A and B in the expressions for $d_{\min}$, the resolution and α_{opt}, the optimum aperture angle given by

$$d_{\min} = BC_s^{1/4}\lambda^{3/4} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\alpha_{\text{opt}} = A(\lambda/C_s)^{1/4} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and since values varying by a factor 2 have been variously determined for these quantities it becomes important to know what is the correct method of making the analysis.

Following a treatment of Conrady (1919) for the criterion for the maximum allowable path difference introduced by spherical aberration, Haine (1954) obtains the values $A = 1{\cdot}4$, $B = 0{\cdot}43$. Using these figures together with the values of C_s and λ appropriate to the Siemens Elmiskop I microscope we obtain a figure of 2·8 Å as the *instrumental* resolving power.

A number of proposals have been made for the correction of spherical aberration and at least one (Scherzer 1947, 1949 a) has been tried experimentally (Seeliger 1953, Möllenstedt 1956) but as yet none has been brought to a sufficient level of practical perfection to be incorporated in a microscope. Haine and Mulvey (1952) following a proposal of Gabor (1949) have developed an alternative system of diffraction microscopy involving the reconstruction of a primary electron image in an optical apparatus designed so as to correct the aberrations of the electron optical system. At present it appears that the electrical and mechanical stability required over long exposure times for the success of such a system cannot yet be readily achieved.

The attainable practical resolution is somewhat worsened from the instrumental resolving power quoted above as a result of astigmatism and chromatic aberration. Astigmatism arises from lack of rotational symmetry in the objective lens fields produced by asymmetries in the machining of the pole pieces or possibly inhomogeneity in the iron from which they are made. As a result of this a point in the object plane is imaged as a circular disc between the stigmatic image points. If the astigmatic focal length difference is Δf, the diameter of this disc in the object plane is

$$\delta_A = \alpha \Delta f. \qquad (3)$$

For this to be negligible compared with the instrumental resolving power, say $\delta_A \sim 0{\cdot}2\, d_{\min}$, i.e. $\Delta f \sim 6 \times 10^{-7}$ cm.

The relative astigmatic current difference $\Delta J_A/J$ should then satisfy the condition

$$\Delta f = \frac{\Delta J_A}{J} \leqslant 1{\cdot}1 \times 10^{-6}. \qquad (4)$$

A test for this condition has been developed by Hillier and Ramberg (1947) and Haine and Mulvey (1954) by observing the asymmetry of the Fresnel fringe around a small hole in a carbon film as the objective lens passes through focus. Stigmators have been incorporated in modern electron microscope objective lenses which introduce a weak elliptical field the strength and direction of which may be varied to compensate the natural astigmatism of the lens. The degree to which this correction can be made exact depends upon visual estimates either on the fluorescent screen or the photographic plate of the asymmetry of the Fresnel fringe. The best adjustment that can readily be made in practice corresponds to $\Delta J_A/J \sim 2 \times 10^{-5}$, so that the condition (4) is hardly satisfied and astigmatism therefore makes a very real contribution to loss of resolution.

If the accelerating voltage U changes by ΔU and the current J in the winding of the objective lens by ΔJ then the focal length of the lens changes to cause a chromatic disc of confusion of diameter

$$\delta_F = 2C_F\alpha\left\{\left(\frac{\Delta U}{U}\right)^2 + \left(\frac{2\Delta J}{J}\right)^2\right\}^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where C_F, the chromatic aberration constant of the lens has the value 2·2 mm for the Siemens microscope. For these to have negligible effect on the resolution

$$\frac{\Delta U}{U} \leqslant 1{\cdot}5 \times 10^{-6} \text{ and } \frac{\Delta J}{J} \leqslant 0{\cdot}75 \times 10^{-6}. \quad . \quad . \quad . \quad . \quad (6)$$

Such stabilities are not readily achieved for the necessary focusing and exposure time without special equipment (Haine and Jervis 1954). There is no practical advantage to be gained however, at the moment, in using very elaborate stabilizing equipment since there is another source of chromatic aberration in the energy losses suffered by inelastic collision of the electrons with the specimen. This may be minimized by using the thinnest possible specimens but cannot be entirely eliminated without the use of some filtering device (this would have to be an electrostatic lens) and even then severe difficulties may arise as a result of loss in brightness of the image. In practice therefore there is some loss of resolution due to chromatic aberration.

If the four aberrations are comparable in magnitude it is again not too easy to combine them together rigorously to compute the resolution. However with reasonable assumptions, it can be confidently stated that a practical resolution, as limited by the factors already discussed, of 5–6 A may be achieved.

2.3. *Resolution and Contrast*

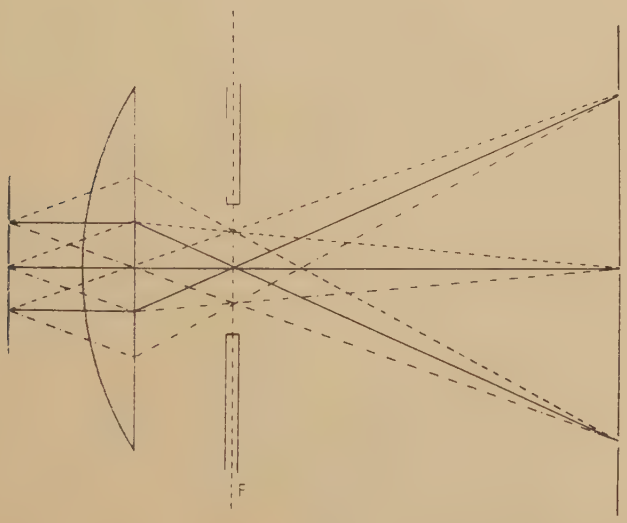
At this level of resolution the form of the specimen itself must also be taken into account in deciding exactly what will be seen in the image. A single unit cell of gold, for example is a cube with sides ~ 4 A long containing fourteen atoms. This type of specimen raises two questions which have not yet received adequate answers. The first is the likely contrast of such an object against a supporting film, and the second is a consideration in detail of the manner in which the specimen scatters the incident beam. The problem of the single atom has been considered by a number of workers (Hillier 1941, Marton and Schiff 1941, Boersch 1947, Haine 1957.) Their conclusions differ slightly depending on assumptions made about the detailed scattering process, but they are all generally agreed that individual atoms at least of higher atomic numbers should be visible in a miscroscope with a resolution of 2–3 A. The experimental realization of a monatomic layer of isolated (i.e. non-interacting) atoms is not readily achieved. The type of specimen we are concerned with in practice may take the form of a monolayer of atoms or molecules on a supporting film. There may well be some special geometrical relationship between the

positions of these atoms. Another type of specimen may be a small spherical aggregate of atoms either within a matrix or attached to a supporting film. Again we may have a monatomic planar aggregate of atoms such as a disc of copper atoms in a Guinier–Preston zone of an Al–Cu alloy. Now that objects of these dimensions are within the range of possible observation there is a serious need for some detailed theoretical consideration of the effects they may be expected to give in electron microscope images.

2.4. *Selected Area Diffraction*

Before concluding this discussion of the salient features of the modern electron microscope mention should be made of a facility which proves particularly useful in the direct examination of crystals. Figure 1 shows diagrammatically the ray paths through a lens imaging a periodic object.

Fig. 1



Ray diagram illustrating action of lens in imaging a periodic object. The diffraction pattern is formed in the back focal plane of the lens and the image in the image plane.

Diffracted beams from the object are focused to a point (image of the source) in the back focal plane of the objective lens. In the electron microscope it is possible to introduce a limiting aperture in the intermediate image plane to restrict to as small as $0{\cdot}2\ \mu^2$ the area contributing to the image. Further, it is possible by reducing the strength of the first projector lens to form on the final screen with the second projector lens an enlarged image of the back focal plane of the objective lens. Under these conditions one thus obtains an image of the diffraction pattern of a highly selected area of the specimen. This facility has a twofold advantage

in the study of crystals. Firstly we may obtain all the information normally extracted from diffraction patterns, and secondly we may more readily interpret many of the contrast effects observed in images of crystals.

§ 3. Surface Methods of Studying Molecular Detail

3.1. *Study of Macromolecular Crystal Lattices by Surface Replica Techniques*

If the molecules of a crystal are sufficiently large then some aspects of their arrangement within the crystal may usefully be studied by observations upon the crystal surface. Replica techniques have been extensively applied by Wyckoff and his co-workers to the study of protein and virus crystals in which the individual macromolecules are so large as to cause an appreciable periodic undulation of the crystal surface which is made visible by suitable replica methods. In the earlier experiments (Price and Wyckoff 1946) a pseudo-replica was made by evaporating a thin palladium shadowing layer obliquely on to the surface of crystals dispersed on a glass slide. This was strengthened by means of a thin backing film of collodion (later replaced by silicon monoxide in a modification of the method due to Hall 1951). The pseudo-replica was detached from the crystal by solution of the latter. In this way the highly symmetrical arrangement of the molecules in the crystal could be directly revealed. Using a similar method Dawson (1951) was able to reveal monomolecular steps and rows of molecules running parallel to the steps on the surface of crystals of β-lactoglobulin (M.W. 36 000). He determined two of the lattice dimensions in this crystal as 105 ∓ 5 Å and 50 ∓ 10 Å in good agreement with the low angle x-ray measurements of Riley (1951). The value of the original method using collodion as a backing film was limited by the difficulty often encountered in dissolving the crystal away from the replica after metal shadowing, the self structure of the collodion, and its inability to withstand electron bombardment at the high intensity required to reveal the molecular detail. A considerable improvement in the quality of these pseudo-replicas came as a result of the use of carbon films, following the method of Bradley (1954), as backing films for the metal shadowing layer. Their great transparency to electrons, freedom from self structure and high mechanical strength have permitted the preparation of durable replicas of extensive areas of crystal surface and simultaneous observation of several crystal faces. In numerous papers from Wyckoff's laboratory in the last few years many examples of molecular arrays have been shown. Crystals of southern bean mosaic virus protein consist of cubic close packed spheres of diameter 245 Å with a cube edge length of 345 Å (Labaw and Wyckoff 1955, 1957 a) fig. 2, Pl. 25). Another plant protein crystal with a molecular size of 140 Å is shown in fig. 3, Pl. 26. The effect of salt content on the crystal structure of turnip yellow virus protein has been studied (Labaw and Wyckoff 1956 a), the intermolecular distance changing from 300 Å to 210 Å on the removal of salt.

These authors claim that detail of the order 20–30 Å may be resolved by this replica method and have recently carried out experiments to determine the useful limit of the method before resorting to the more direct methods of lattice imaging developed by Menter (1956 a). Successful replicas showing two dimensional surface arrays of molecules have been obtained from concanavallin B (M.W. 42 000) with a lattice size on the exposed face of 62 Å × 87 Å (Labaw and Wyckoff 1956 b). The image consists of an array of dots of average diameter 30–40 Å, the dots being physically separated one from another. The authors suggest that the limiting factor is not so much the faithfulness of the replica being marred by the granularity of the shadowing metal, as changes in the specimen on electron bombardment and debris left on the crystal surface obscuring the surface detail. The latter is thought to consist often of very small protein molecules adsorbed on the larger molecules and their obscuring effect becomes more important as the size of the molecules comprising the crystal is reduced. In some earlier photographs of larger molecular lattices there is indeed apparent a suggestion of fine identical detail on adjacent molecules on the surface. Wyckoff believes that this does not represent true detail of the molecule itself but is probably due to adsorption of the smaller protein molecules at preferred sites on the surface of the larger molecule.

Throughout this work Wyckoff has commented on the striking absence of evidence for dislocations in these crystals.

Only rarely have spiral terraces corresponding to screw dislocations been observed, suggesting that the Frank mechanism may not be essential for the growth of these crystals.

3.2. *Molecular Detail Revealed by Direct Examination of Shadowed Crystals*

With the virus and protein crystals it is necessary to use a replica method on account of the comparatively great thickness of the crystals. However, with plate-like crystals it has been found possible to study molecular surface detail upon them by direct examination of shadowed preparations of the crystals themselves. This method has been particularly successful in the study of the growth mechanism of crystals of straight chain hydrocarbons and their derivatives by Dawson and co-workers. (Dawson and Vand 1951 a, b, Dawson 1952, Anderson and Dawson 1953, 1955). In the first series of experiments crystals of n-hexatriacontane, n–$C_{36}H_{74}$, were prepared by crystallizing directly from solution on a supporting film. Oblique shadowing revealed the presence of spiral growth terraces of height 43 ∓ 5 Å, a distance close to the length of the hydrocarbon chain measured from x-ray diffraction photographs, 47·51 Å. These steps are therefore monomolecular and their spiral form confirmed in a most striking manner the growth mechanism postulated by Burton *et al.* (1951) based on the presence of a screw dislocation in the crystal with its Burgers vector normal to the growth surface. Many other

phenomena predicted by the theory such as the observation of complete layers of monomolecular step height arising from growth centred on two neighbouring dislocations of opposite hand were also observed. These observations have been adequately reviewed by Forty (1954). This work has been extended in several directions. A comparative study (Anderson and Dawson 1953) was made of step heights on n-nonatriacontane, $C_{39}H_{80}$ with a chain length of 51·3 Å and stearic acid, $C_{17}H_{35}COOH$ with a chain length of 25 Å. Crystals of the former grow with step heights of 45 ∓ 10 Å and the latter with step heights of 47 ∓ 10 Å confirming that the height is determined by the state of association of the molecules in the liquid phase, the stearic acid molecules being associated into dimers on account of the carboxyl dipole at the end of the molecule. Using the ester n-propyl n-pentacontanoate, $n\text{–}C_{53}H_{106}O_{21}$ evidence has been obtained for monomolecular sheet nuclei $0{\cdot}25\text{–}20\mu^2$ in area and for the formation of screw dislocations as a result of the incomplete lattice register of molecules adsorbed at the edges of these sheets. The authors suggest that these observations cast doubt on the earlier proposal of Frank (1951) that the screw dislocation promoting growth is introduced into the monomolecular nucleus as a result of a buckling mechanism.

Gard *et al.* (1955) have used the direct shadowing method to study growth steps on potash felspar crystals. Figure 4, Pl. 27 shows a crystal shadowed with Au/Pd at an angle $\tan^{-1}\frac{1}{4}$ on which a series of parallel steps may be seen. Gard estimates from the minimum shadow length resolvable in his instrument that the steps are probably two unit cells high (~ 26 Å) and may possibly be only one cell high.

3.2.1. *Inadequacy of Shadowing*

It is clear from fig. 4 that even using an instrument with better resolution than Gard's estimated 50 Å this method of revealing monomolecular terraces is likely to be limited by the irregularity of the shadow edge arising from the aggregated structure of the evaporated metal layer. This indeed is one of the present most serious problems in electron microscopy. Workers vary widely in their recommendation of the best metal to use for shadowing to give high contrast and least self structure, but there seems to have been no critical comparative study of materials at the highest resolution. Among materials recommended are palladium (Wyckoff 1956), praesodymium or its oxide (Locquin 1956), tungsten oxide (Le Poole 1956) platinum, to mention but a few. The tendency to aggregation and grain growth of the evaporated metal depend markedly on conditions of evaporation and the nature of the substrate onto which the material is deposited. A critical study of the problem would be invaluable. Some attempts have been made to evaporate simultaneously a mixture of a non-aggregating material such as carbon with a heavy metal to give the necessary contrast (Koch and König 1954). Bradley (1958) has recently obtained promising results using the carbon evaporation technique with fabricated rods of carbon containing 80 wt% platinum metal. The shadowing layer appears

to have a remarkably small self structure and good contrast due to the high platinum content. The studies of Keith (1956) on the effect of temperature of the substrate and the nature of the residual atmosphere in the vacuum chamber on grain growth in evaporated copper layers suggests another possible line of approach. He found that evaporated copper layers deposited on to substrates cooled to liquid air temperature showed extensive grain growth from about 30 Å to 200 Å on annealing to room temperature in the absence of oxygen. If, however, oxygen was admitted while the deposit was still cold and the latter was then annealed to room temperature, grain growth was severely inhibited and the resultant crystal size was only 60 Å. It is possible that by using a metal such as tungsten in this way one could obtain a shadowing layer with very small self structure and the necessary contrast.

3.3. *Decoration Technique for Revealing Surface Detail of Atomic Dimensions*

A new approach to the study of certain types of surface geometry down to atomic dimensions has recently been made by Bassett (1958) working in the author's laboratory. A small quantity of gold (corresponding to an average thickness of 3–5 Å if spread uniformly over the surface) is evaporated *in vacuo* on to a cleavage surface of rocksalt or lithium fluoride. The gold deposits on the surface have discrete nuclei of two types. The first, from 60 Å downwards in size appear to be dispersed randomly over the crystal surface. The second are rather smaller (25 Å downwards) and form preferentially in large numbers along surface steps. The distribution of nuclei may be studied unaltered in the electron microscope by evaporating a carbon film on to the crystal and removing this with the nuclei attached to it by dissolving the crystal. Figure 5, Pl. 27 shows a typical array of cleavage steps decorated by this new technique. The method is a considerable improvement over earlier decoration techniques used with the optical microscope since it has a very high resolution in the plane of the surface. Furthermore, although it is not possible to measure directly the height of a single monatomic step it may be proved conclusively that steps of this height are decorated. Figure 6, Pl. 28 shows a terrace of cleavage steps crossed at right angles by a straight step which is displaced sideways by about 150 Å as it crosses the terrace. The straight step is thought to arise by slip on one of the (110) planes inclined at 45° to the cleavage surface. In crossing a cleavage terrace the line of intersection of this slip plane with the cleavage surface is displaced in a direction normal to itself by an amount equal to the height of the terrace. The individual steps comprising this terrace may be counted (45 ∓ 2) and if the height of each is a half unit cell, i.e. 2·8 Å, the expected total height of the terrace is 125 ∓ 6 Å. Thus most of the steps in this terrace must be elementary lattice steps.

Bassett has shown that deliberately produced slip steps on lithium fluoride may be decorated in this way. It is early yet to say how general

the technique may be for studying detail of atomic dimensions on surfaces. Little is known about the conditions for surface mobility of the deposited metal, a problem of great intrinsic interest on which Bassett's experiments may throw some light. It is to be expected that similar phenomena may be observed with electrodeposited layers. Vermilyea (private communication) has found an unusually high over-potential for electrodeposition on certain whiskers which he associates with the absence of surface steps for nucleation. When electrodeposition does occur on an imperfect surface one may therefore expect that nuclei will form preferentially along steps as in evaporation.

3.4. *Direct Methods of Studying Surface Structure*

Even with improvements in resolving power that might be derived from advances in replica preparation there are still serious drawbacks inherent in the technique which limit its usefulness. The replica provides information only about the surface of the crystal from which information about the interior must be inferred. Furthermore the replica is effectively only a single point observation on the crystal so that it is virtually impossible to follow any changes in the same specimen resulting from changes in parameters such as temperature or stress. Interest has therefore moved increasingly towards direct techniques of examination in which the crystal itself constitutes the specimen. These may be classified into two groups, namely, methods in which the surface of the crystal is examined directly either by making it self emitting or by scattering electrons or ions from the surface, and transmission methods in which an electron beam passes through a thin crystal. Among the former methods are the field emission microscope, particularly the ion form developed by Müller (1951, 1956, 1957), the electron scanning microscope (a recent form has been described by Macmullan 1952 and Smith and Oatley 1955), the reflection electron microscope, the secondary emission microscope (Möllenstedt and Düker 1953). As with replica methods only the surface of the specimen is accessible for observation but direct observation of changes does become possible. Apart from the field emission methods, the resolving power of these methods is comparatively poor and only in exceptionally favourable circumstances can any of them be used for studying detail of molecular dimensions. Menter (1952, 1953) has shown that sharp steps on surfaces associated either with growth or cleavage have exceptionally high contrast in the reflection electron microscope. Cleavage steps less than 100 A high on the surface of mica, and growth terraces associated with screw dislocations of Burgers vector possibly one or two unit cells long in silicon carbide crystals, may be detected. It is not easy to measure these step heights accurately on account of lack of knowledge of the detailed electron scattering processes in such features under the illuminating conditions obtaining in the reflection microscope, so that quantitative work is hardly possible.

The field ion microscope certainly provides a remarkably good resolution. Müller (1957) has obtained photographs of tungsten points showing the atomic array with neighbouring atoms less than 3 Å apart and some disordered features have been interpreted in terms of dislocations emerging from the specimen surface (see also Drechsler 1956, Drechsler *et al.* 1955). As a general method for studying lattice phenomena, however, the technique has a number of serious disadvantages. The specimen must be produced in the form of a very fine point with a radius at the tip of the order of 10^{-5} cm or less, the optimum resolution is only obtained in the neighbourhood of liquid hydrogen temperatures, and the specimen is subjected to very high mechanical stress ($\sim 10^6$ lb/in^2) during examination on account of the high field near the tip. For the direct study of surface phenomena at the atomic level the technique is unsurpassed. For details of this work reference should be made to the review of Good and Müller (1956).

§ 4. Observation of Crystal Lattices by Direct Transmission

4.1. *Macromolecular Crystals*

With the development of microtomes capable of cutting sections below 200 Å in thickness it became possible to study the structure of macromolecular lattices of virus and protein crystals by direct transmission. Farrant and Hodge (1956) succeeded in cutting 200 Å sections of crystals of the protein ferritin. High contrast is obtained in these images on account of an unusually high concentration of heavy atoms in the form of a micelle of ferric hydroxide some 55 Å in diameter in each molecule, and some micrographs clearly reveal a square array of molecules about 80 Å apart. Bergold and Fernandez-Moran (private communication) have obtained similar photographs of sections of polyhedral protein crystals formed in insect cells at a late stage in virus multiplication (fig. 7, Pl. 29). Both authors have remarked on the absence of evidence for dislocations in these crystals in line with the similar evidence obtained from surface replica studies. In the latter case this is the more remarkable since the crystals, in growing, occlude the viruses which can be seen as irregularly shaped bodies in the section, with no deleterious effect on the perfection of the surrounding protein crystal lattice.

Farrant and Hodge describe a contrast reversal effect observed in out-of-focus images of their crystal sections similar to the optical effect seen in off-focus images of periodic objects such as gratings. This suggests that the crystal section is acting as a true grating for electrons and that there is some definite phase relationship between electrons scattered from neighbouring molecules. Evidence for the production of well defined diffracted beams from macromolecular crystal sections has recently been obtained by Karasaki and Komoda (1958). These authors observed fringe systems with a spacing of 70 Å in sections of yolk platelets included in certain embryonic cells. These have been interpreted as arising from

net-planes in the macromolecular crystal since diffraction spots corresponding to a spacing of 72 A are observed in the back focal plane of the objective. It appears then that in some cases images of macromolecular organic crystals are essentially Abbe images similar to those obtained from much smaller molecules to be described later. This cannot always be so, since many organic molecules, particularly those of biological importance are known to be degraded by electron irradiation. Once this has occurred then the exact spatial relationship between corresponding atoms from one molecule to the next required for the formation of diffracted beams is lost and contrast can only arise as a result of the periodic variation in mass thickness from the centre of one molecule to the next.

A striking feature of these sections of macromolecular crystals is the apparent freedom from gross distortion introduced as a result of the cutting action of the microtome knife. There is hardly enough evidence as yet to decide whether this applies equally to materials of physical and metallurgical interest. Using a polished diamond knife, Fernandez-Moran (1956) and Haanstra (1955) have shown that it is possible to cut thin sections of a number of metals but there has been no critical evaluation of the plastic deformation introduced by the cutting.

4.2. *Crystals with Small Lattice Parameters*

With the recent development of very high resolution electron microscope it has now become possible to image lattices by transmission of the electron beam through crystals, with net plane spacings down to about 5 A. The early theoretical considerations of Boersch (1947) already showed that crystal lattice images should be obtainable in this way, essentially by recombination of diffraction spectra from the crystal, by the objective lens of the microscope.

4.2.1. *General conditions for lattice resolution*

The first images of this type were obtained by Menter (1956 a). He selected the phthalocyanine class of compounds to study on account of their relatively high lattice parameters, some of which were comparable in magnitude with the stated resolution limit for two point objects of the electron microscope available (~ 10 A). A number of further conditions essential for lattice imaging were also fulfilled by this class of compounds. These include :

(1) Suitable habit so that when prepared in the normal way with the crystals lying on a horizontal supporting film, there is a reasonable probability of finding crystals with the lattice planes to be resolved lying closely normal to the supporting film. They are then in a favourable orientation to produce the essential diffraction spectra for image formation.

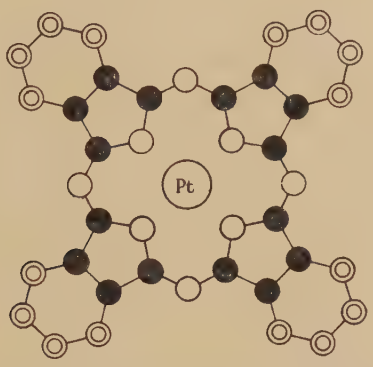
(2) The structure factor for the reflections from the lattice planes to be resolved must be reasonably high so that there is adequate contrast in the image.

(3) The crystals must be resistant to chemical degradation by electron bombardment.

(4) The crystals must be thermally stable since they may in some circumstances suffer an appreciable temperature rise during observation.

(5) The crystals must be sufficiently thin in the direction of the electron beam in order to reduce to a minimum effects of inelastic scattering.

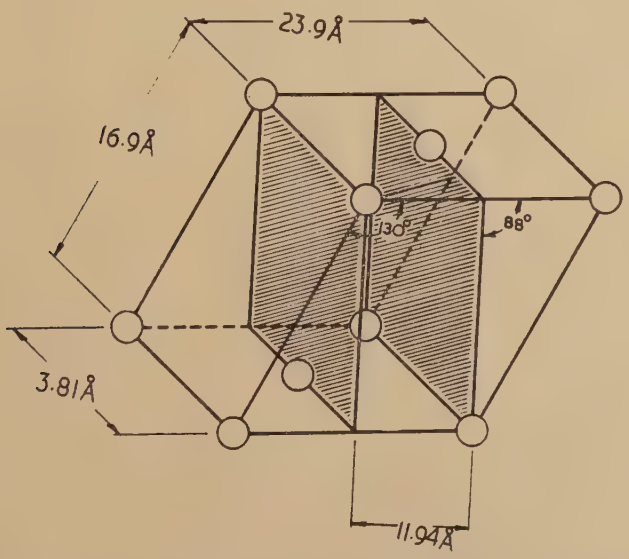
Fig. 8



From *Proc. roy. Soc. A.*

Schematic illustration of platinum phthalocyanine molecule, ●, C; ◎, CH: ○, N.. Robertson and Woodward (1940).

Fig. 9



From *Proc. roy. Soc. A.*

Unit cell of platinum phthalocyanine crystal. Circles represent molecules, the metal atom being situated at the centre of the circle. Robertson and Woodward (1940).

Menter worked mainly with the simple metal derivatives of phthalocyanine, obtained by replacing two hydrogen atoms at the centre of the molecule with the metal atom. Figure 8 shows the platinum phthalocyanine molecule and fig. 9 the unit cell of this compound. It is monoclinic with $a_0 = 23{\cdot}9$ Å, $b_0 = 3{\cdot}81$ Å, $c_0 = 16{\cdot}9$ Å, $\beta = 129{\cdot}6°$ (Robertson and Woodward 1940) and frequently exhibits a ribbon like habit with elongation along b_0, the ribbon surface being (001). Thus when deposited from a liquid suspension on to a supporting film the crystals tend to settle on (001). Now the angle (001) (20$\bar{1}$) is $\sim 88°$ so that the (20$\bar{1}$) planes are then almost parallel with the electron beam and are therefore in a suitable orientation for diffraction since the Bragg angles are very small for electrons. Figure 10, Pl. 30 shows an electron micrograph of a platinum phthalocyanine crystal, a few hundred Ångstrom units in thickness. The spacing of the parallel lines is 12·0 Å with a standard deviation of 0·2 Å, which is consistent with the x-ray value 11·94 Å (Robertson and Woodward 1940) for the spacing of the (20$\bar{1}$) planes. The exact significance of such an image may be understood in terms of the Abbe theory of the miscroscope (Abbe 1837).

4.2.2. *Diffraction theory of image formation*

Figure 1 shows the ray paths through a lens imaging an object consisting of a bar and space grating with elementary spacing d. The incident radiation is analysed by the grating into a series of diffracted beams. The ray paths for the zero order and first order spectra only are shown. The function of the lens is to recombine these spectra with the appropriate phases to form an image of the grating in the image plane of the lens. If the incident illumination has a very small divergence then the diffracted spectra are near parallel beams and are therefore focused in the back focal plane F of the lens, so that a diffraction pattern is formed in this plane. We thus have two complementary sets of data about the grating, the diffraction pattern in the back focal plane and the image in the image plane. Since the aperture of the lens, particularly an electron lens, must be severely restricted in order to optimize the effects of aberrations, the number of diffracted beams contributing to the image is limited. On this account, the image formed by a lens is always imperfect, and as the aperture is reduced the image becomes a more and more crude representation of the object. Indeed when there is a small number of spectra, false detail may appear in the image which shows up in a particularly striking way with periodic objects (Porter 1906). However, in order to reveal the basic periodicity of a grating, the necessary and sufficient condition is that the aperture of the system should be such as to pass at least the zero order spectrum together with one of the first order spectra. With axial illumination and axially symmetrical optical systems as commonly used this generally means in practice the inclusion of both first order spectra on either side of the zero order. Thus in order to detect a periodic line structure of spacing d it is necessary that the semi-aperture α shall be greater than $\lambda/2d$ where λ is the wavelength of the

radiation used. This may be simply demonstrated optically by imaging a square grid (fig. 11 (*a*), Pl. 30. The diffraction pattern of this object recorded in the back focal plane is shown in fig. 11 (*c*), Pl. 30. It consists of a square array of spots (doubled because of the two wavelengths in the filtered mercury source). If the aperture is reduced so that only S_0 and the first order spectra S_1, S_2 contribute to the image the grid bars giving rise to the spectra S_3, S_4 disappear from the image (fig. 11 (*b*), Pl. 30.

These considerations may be applied directly to the imaging of lattices by electron lenses. The crystal being thin in the direction of the electron beam produces a cross grating diffraction pattern in the back focal plane of the objective lens (figure 12, Pl. 30). With platinum phthalocyanine in the orientation studied this pattern is rectangular, the unit dimensions of which correspond to those obtained from the mutually perpendicular net planes (020) ($d = 1{\cdot}95$ Å) and ($20\bar{1}$) ($d = 11{\cdot}94$ Å). A 30 μ aperture superimposed on this pattern permits only the zero and first order spectra of the ($20\bar{1}$) planes to pass to the image plane where they recombine to give a periodic line structure of spacing 11·94 Å. Thus the image is essentially a very poor Fourier projection of the crystal structure in the direction of the electron beam. The image can be said to reveal one aspect of the periodic arrangement of the molecules in the crystal. The validity of this condition for the formation of the image has been tested directly by deliberately reducing the objective aperture until the first order spectra are excluded from the image. Figure 13 (*a*), Pl. 31 shows a crystal of platinum phthalocyanine imaged under these conditions (objective aperture 10 μ) compared with fig. 13 (*b*), Pl. 31 showing the same area observed with an objective aperture of 30 μ. The lattice is not resolved in the former and it is interesting to note that the resolution of non-periodic detail in this image is very poor because the aperture used is some five times smaller than the optimum for imaging such detail.

It might be thought that improvement in the detail of a periodic object could be achieved by increasing the size of the objective aperture to permit higher order spectra to contribute to the image. This may be true for structures of relatively large spacing (see below) but for the smaller spacings the effects of spherical aberration must be considered. The effect of spherical aberration is to impose a phase delay $\epsilon = \frac{1}{4}C_s\alpha^4$ upon a ray passing through the lens at an angle α where C_s is the spherical aberration coefficient and ϵ the linear path difference between this ray and an axial ray. In crystalline materials where the atoms or molecules are arrayed in a regular lattice, there is phase coherence between waves scattered from all the atoms, and strong diffracted beams are transmitted in special directions as narrow pencils by a small specimen. These beams, the relevant ones from the point of view of imaging the periodic structure pass through a very narrow zone of the lens. Under these conditions, as has been pointed out by Scherzer (1949 b), the effects of spherical aberration are less deleterious. Assuming a parallel illuminating beam the phase difference $\Delta\epsilon$, between the two edges of a wave front of width Δr passing

through the lens at a distance r from the axis ($r \sim f\alpha$, where f is the focal length of the lens) is given by $C_s r^3 \Delta r / f^4$. In the case of a platinum phthalocyanine crystal 1000 Å wide this gives $\Delta\epsilon \sim 3 \times 10^{-13}$ cm, using values of $C_s = 0{\cdot}28$ cm, $f = 0{\cdot}3$ cm, for the $20\bar{1}$ reflection. Now $\lambda = 4 \times 10^{-10}$ cm so that $\Delta\epsilon << \lambda$ and the wavefront is sufficiently undistorted to remain coherent with the axial wavefront. It is therefore able to interfere with the latter to form a periodic image. There is, of course, a phase delay of $\frac{1}{4}C_s\alpha^4$ imposed upon the wavefront as a whole which leads to a path difference of $\lambda/40$ but this can be restored by adjusting the focal setting of the lens. If the lens is defocused by an amount Δf then an additional phase shift of $\pi\Delta f\alpha^2/\lambda$ is introduced into the diffracted beam so that the correct phase relationship may be restored by making the appropriate amount of defocusing. In practice the observer may often do this unconsciously since in focusing for the best image of the lattice planes one may often observe that the non-periodic detail of the image is some distance from true focus as shown by the Fresnel fringes along edges (see for example fig. 14, Pl. 32).

The phase difference $\Delta\epsilon$ between the extremities of the wavefront increases rapidly as α increases and this may determine whether any detail is gained by adding higher order spectra to the image. It has not been possible to detect any significant difference between the detail of the periodic structure in images of platinum phthalocyanine crystals with $(40\bar{2})$ and $(\bar{4}02)$ spectra added to the image. On the other hand in crystals of faujasite (cubic, $a_0 = 25$ Å) where (111) planes with a spacing of 14·4 Å are resolved and the second order spectrum is included in the image there does appear in some cases to be slight evidence of a fine structure in the lines of the image (fig. 14). Brindley *et al.* (1958) have observed a superlattice spacing of ~ 100 Å in crystals of the mineral antigorite where a sharpening of the lines results it is thought from the inclusion of higher order spectra in the image (see §4.3.5.). This would certainly be expected since many orders of spectra from such a grating can certainly be recombined even by a relatively poor lens without serious loss of phase coherence. The experiment with platinum phthalocyanine is not conclusive since it is possible that the second order diffraction spectrum observed in the diffraction pattern may not necessarily have originated from a part of the crystal also reflecting in the first order. In conducting experiments of this sort it would be essential to confirm this point by observations on dark field images.

4.2.3. *Limits of orientation for lattice resolution*

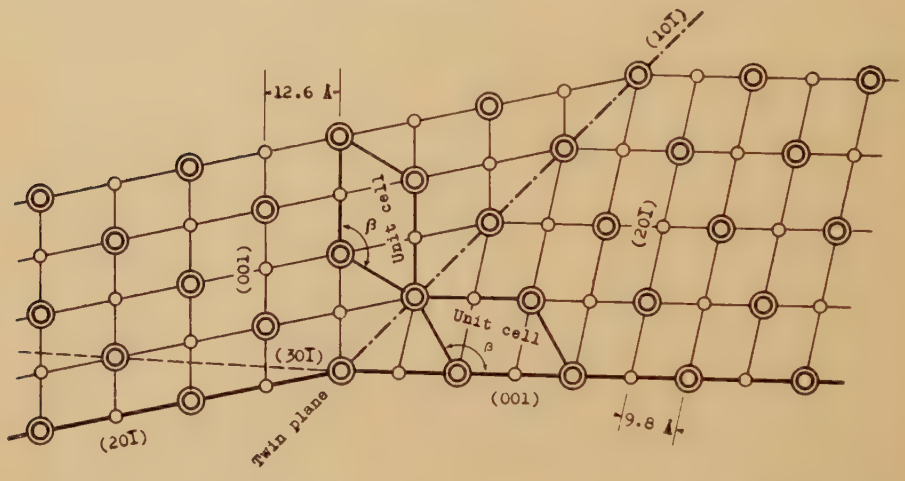
Among the conditions required for lattice imaging the most important is clearly that requiring the lattice planes to be appropriately oriented for the production of diffracted beams. It is therefore of practical importance to consider how far the angular setting may deviate from the ideal. Crystals of platy habit are frequently not flat and this gives rise to the appearance of characteristic extinction contours. The origin

of these contours is clear from fig. 1. We suppose that the aperture size is such as to interrupt the passage of all the diffracted beams from the crystal. Those parts of the crystal from which the diffraction spectra originate will be deficient in electrons and appear ' opaque '. These regions are called Bragg or extinction contours. Any particular contour defines a region of the crystal in which the lattice planes are appropriately oriented to cause strong diffraction of the electrons into one or more orders of the spectrum. Figure 13 (*a*) shows such a contour in a bent crystal of platinum phthalocyanine imaged with a very small objective aperture suitable for excluding all spectra other than the zero order from the image. If the aperture is enlarged so as to permit the first order ($20\bar{1}$) spectrum to contribute to the image then it is found that the lattice planes are resolved in the image of the crystal only in those regions where the extinction contour was observed with the smaller aperture (fig. 13 (*b*). Thus a lattice image is obtained only in regions of the crystal which are diffracting, and in order to determine when the lattice image may be expected to be observed it is necessary to consider in more detail the relation between the reflecting limits and the angular setting of the crystal with respect to the electron beam.

It is well known from electron diffraction experiments that the so-called third Laue condition for diffraction by atom rows in a direction normal to the surface of a thin crystalline specimen need not be rigorously satisfied. To a first approximation this statement may be taken as equivalent to stating that lattice points in reciprocal space are elongated into lines normal to the specimen surface. In the usual reciprocal lattice construction it is assumed that Bragg reflections will be obtained from net planes when the sphere of reflection passes through the reciprocal lattice point. The elongation of the point into a line means therefore that there is some latitude allowed in the angular setting of the crystal and the net planes corresponding to this elongated point may reflect over an appreciable angle. Using kinematical theory Menter (1956 a) has, on this basis, suggested that with the ($20\bar{1}$) planes in a platinum phthalocyanine crystal 200 Å in thickness the permitted misorientation is $\sim 6°$. Some supporting evidence for this considerable relaxation of the diffraction condition was obtained from a ribbon like crystal apparently twisted into a helical form (fig. 15, Pl. 33). The lattice planes are resolved for an appreciable distance along the direction of the axis of the helix, at every half twist in the crystal. Assuming that the crystal is uniformly twisted we may deduce that the lattice is twisted away from the optimum orientation by $\sim 6°$ at the extremities of the regions where it is resolved in the image. Wyckoff (private communication) has observed a similar effect in crystals of indanthrene scarlet. The adequacy of a kinematic treatment of this problem has been questioned by Neider (1957), who suggests that the dynamical theory of electron diffraction must be used for a full explanation of crystal lattice images (see Niehrs 1956). The two main consequences of a dynamical treatment are that the contrast of the image will vary

periodically with thickness due to extinction and that the angular range of reflection is severely restricted. With certain simplifying assumptions Neider calculates that at a deviation of 2·5° from the Bragg angle for a crystal of nickel phthalocyanine of optimum thickness, the scattered intensity is reduced to less than 4% of the maximum value and at 10° to 0·1%. Experimental evidence is advanced to support this view based upon observations on the disappearance of the lattice image as a result of tilting the specimen away from the optimum orientation. While it is likely that Neider's conclusions are qualitatively correct, at least for rather thick crystals, it is difficult to assess their quantitative validity, when so many simplifying assumptions are required to make a calculation of the expected intensity distribution.

Fig. 17



From *Nature, Lond*

Proposed twin configuration in copper phthalocyanine to account for image seen in fig. 17. In the electron microscope this structure (here projected along b_0) is seen as viewed from top of page. Small circles are at level $\mp \frac{1}{2}b_0$. Komoda *et al.* (1958).

Some supporting evidence that the third Laue condition may sometimes be highly relaxed comes from recent observations of Komoda *et al.* (1958) on twinned crystals of copper phthalocyanine. Figure 16, Pl. 34 shows a crystal in which the spacing of the resolved lattice changes abruptly across the line AB, from 12·5 Å corresponding to (001) lattice planes to 9·7 Å corresponding to (20$\bar{1}$) planes. The authors have interpreted this as due to a twin boundary as shown in fig. 17. Now the angle between the net planes resolved on either side of this boundary is 14° and it must be supposed that unless the crystal is severely bent, the third Laue condition must be highly relaxed. This may well be the case, for unless the crystal

were extremely thin, one would expect an appreciable breadth of disturbed region at the junction of the two lattices because of the obliquity of the twin plane with respect to the surface.

4.2.4. *Limitation of kinematic theory*

The diffraction theory for image formation given above is based on the kinematic theory of electron diffraction, which gives only a geometrical description of the diffraction pattern, assuming no interaction between diffracted waves inside the crystal. In reality, strong diffracted waves are formed already inside the crystal and as a result of interactions between these and the zero order beam inside the crystal, the actual intensity distribution of electron current on the exit surface of the crystal may be considerably modified (dynamical theory). As a major consequence of this the amount of energy appearing in a given diffracted beam varies periodically with crystal thickness (neglecting absorption). The formation of extinction contours in the form of striations parallel to the edge of a wedge-shaped crystal as a result of this effect is well known (see Heidenreich 1942, Kinder 1943, Kossel 1943). Niehrs (1956) has been able to account quantitatively for such effects seen in polyhedral crystals of magnesium oxide. More recently, particularly beautiful examples have been found associated with stacking faults in thin films of stainless steel by Whelan *et al.* (1957) and explained in detail by Whelan and Hirsch (1957 a, b). Striations of this type are generally spaced by several hundred Å with the wedge angles normally encountered and may therefore be distinguished from the more closely spaced lines associated with lattice resolution. It can be seen, however, that extinction effects will certainly modify the contrast of lattice images as pointed out by Niehrs (1957), determining the absolute level of contrast of the image in parallel sided crystals and giving a long period modulation of the intensity in wedge-shaped crystals where the resolved lattice planes lie parallel to the apex of the wedge. Hashimoto (private communication) has pointed out that in a situation such as may occur in fig. 14 where the thickness of the crystal decreases from left to right and the resolved lattice planes are not parallel to contours of equal thickness in the wedge, then from the theory of Niehrs (1954) we may expect a stepped structure in the image of the planes. There is a periodic reversal of contrast in the lines of the lattice image at each half integral value of the extinction distance. A similar contrast reversal effect has been observed by Whelan and Hirsch (1957 a, b) where interference fringes in stacking faults in stainless steel cross extinction contours in wedge-shaped crystals. It is clear that some caution is required in interpretation of the periodic images of crystals described above as true direct images of lattice planes. In parallel sided crystals there is no difficulty but with crystals of varying thickness pseudo-structures may appear. In both cases however, where extra terminating half lines are seen these may be unequivocally associated with dislocations by making the appropriate Burgers circuit around them (see § 4.2.5.). With extended

dislocations in orientations where the total Burgers is normal to the plane of the specimen, no extra half lines are seen (§ 5.3.2) but a deflection of the image lines occurs across the stacking fault between the partial dislocations. In this particular case, this effect may be indistinguishable from the stepped structure described above.

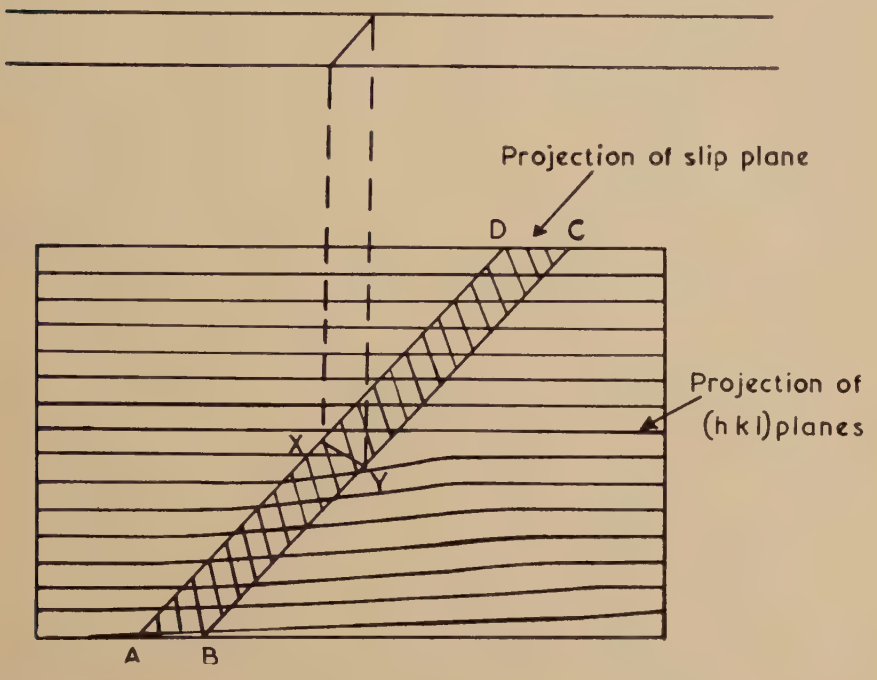
4.2.5. *Dislocations in directly resolved lattices and the significance of 'extra' half-planes*

Menter (1956 a) observed that the perfection of the resolved lattice of platinum phthalocyanine crystals was sometimes disturbed. In a few instances, e.g. figs. 18 and 19, Pl. 34 the disturbance has a particularly simple form associated with an apparent 'extra' half-plane of molecules terminating in the interior of the crystal. It was originally assumed that this could be interpreted as a dislocation line passing through the crystal normal to the flat habit surface with Burgers vector normal to the lattice planes resolved. Further consideration shows, however, that the situation is not quite so simple. A detailed study of the relative orientations of the molecules in adjacent (201) planes in this structure shows on packing grounds alone that it is impossible to insert a single 'extra' half-plane in this way. More detailed consideration of the problem by Bassett *et al.* (1958), shows that the effect is certainly attributable to a dislocation but that the restriction originally imposed upon on the direction of the line and its Burgers vector was incorrect.

Suppose we have a thin crystal with a slip plane $ABCD$ inclined to the surface and a dislocation line XY with Burgers vector $[uvw]$ lying in this slip plane as shown in fig. 20. The effect of introducing this dislocation into the crystal is to shift any set of lattice planes (hkl) on one side of the slip plane with respect to the same set of planes on the other side by the resolved component of the Burgers vector of the dislocation in a direction perpendicular to the (hkl) planes. For a cubic lattice it is readily shown that this distance is given by $d_{hkl}(hu+kv+lw)$ where d_{hkl} is the spacing of the (hkl) planes. Thus the effect of the introduction of the dislocation is to produce $(hu+kv+lw)$ extra inter-planar spacing on one side of the slip plane. If the (hkl) planes are normal to the plane of the specimen, the electron microscope image is essentially a projection of these planes along the incident beam. It follows from the analysis that the 'extra' half-planes seen in such a projection need not be the actual half-planes normally associated for example with an edge dislocation. In the face-centred cubic system there are two 'extra' (220) half-planes but depending upon the orientation of the Burgers vector with respect to the plane of projection we may see 0, 1 or 2 'extra' half-planes in the image. Furthermore the analysis makes assumptions neither about the direction of the dislocation line nor its orientation with respect to its Burgers vector except that it passes through the crystal from one surface to the other. The 'extra' half line or lines are expected to terminate in the vicinity of the projection of the dislocation line on the plane of the image but

need not be uniquely defined in that the disregister associated with them may extend over several lattice spacings in the vicinity of the line. It follows also that a screw dislocation will also give rise to extra half lines provided its Burgers vector is appropriately oriented. Since we do not know the Burgers vector for the phthalocyanine lattice it is not possible to analyse figs. 18 and 19 in detail. It can be shown that the analysis

Fig. 20



From *Proc. roy. Soc. A.*

The dislocation line XY lies in the slip plane ABCD. The effect of introducing the dislocation into the crystal is to shift the right-hand half with respect to the left by an amount equal to the Burgers vector of the dislocation. Planes (*hkl*) are shifted in a direction normal to themselves by an amount equal to the component of the Burgers vector in this direction.

of the number of 'extra' half-planes associated with dislocations in the cubic system is applicable to all lattices so that we may conclude that the effects seen in the phthalocyanine crystals are due to dislocations, but without further work these dislocations cannot be fully described. The difficulty of fitting in an ' extra ' half-plane in this particular orientation is now removed, since we are able to conclude that the total Burgers vector of the dislocation cannot lie in the plane of projection of the image.

4.3. *Information from Crystal Lattice Studies*

Phthalocyanines

Resolved lattices of phthalocyanines have now been observed by a number of workers, e.g. copper (Menter 1956 a; Labaw and Wyckoff 1956 b, Suito and Uyeda 1957), platinum (Menter 1956 a), nickel (Neider 1957), unsubstituted (Menter 1956 a, Labaw and Wyckoff 1956 b). The observation of the resolved lattice in phthalocyanine itself is of special interest since it demonstrates that the presence of heavy metal atoms is not essential for good contrast in the image. It is likely, however, that observation is eased when heavy atoms are present. Provided the structure factor for the appropriate reflection is not seriously reduced by the arrangement of the metal atoms in the crystal, the absolute intensity of electrons scattered by a crystal with heavy metal atoms is larger for a given thickness of crystal than that from a crystal of the same thickness without metal atoms. Thus an equal intensity in the image forming beams may be obtained from a thinner crystal with heavy metal atoms so that the third Laue condition is more highly relaxed and the probability of finding a crystal suitably oriented for diffraction is increased.

Suito and Uyeda (1957) have studied habit modification in copper phthalocyanine. On boiling a suspension of α Cu-phthalocyanine (an orthorhombic modification) in various organic solvents, e.g. toluene or pyridine, they found that it was transformed to the β-monoclinic form which appeared in the form of very thin laths. Selected area diffraction confirmed that most of these crystals exhibited the habit reported by Robertson (1934) with (001) as the most prominent surface of the lath. However, a few crystals showed both by selected area diffraction and by direct lattice resolution that the prominent surface parallel to the supporting film may sometimes be $(20\bar{1})$. This brings the (001) planes almost parallel to the electron beam with a resolved spacing of 12·6 Å. This different habit accounts for the observation of Menter (1956 a) that two spacings around 10 Å and 13 Å were found in images of these crystals, the former corresponding to the $(20\bar{1})$ spacing and the latter to (001).

Hamm and van Norman (1948) had earlier reported a further transformation to occur as a result of sublimation and recrystallization of copper phthalocyanine on the supporting film under the influence of the incident electron beam. Suito and Uyeda find a resolved lattice spacing of 11·9 Å in this modification in close agreement with a value of 11·89 Å deduced from diffraction photographs and suggest that this modification is closely similar to the orthorhombic α-form.

In view of these various possible modifications it is important to exercise care in using the resolved lattice image of copper phthalocyanine for calibrating the magnification of an electron microscope. The orientation and crystal form can generally be checked by observing the diffraction pattern but it is probably safer to use platinum phthalocyanine which does not appear to exhibit this variability of crystal habit and structure.

The unusual habit of copper phthalocyanine has also been observed by Neider (1957) who reports a value of 12·12 Å for the (001) spacing determined from diffraction patterns and resolved on micrographs. A further type of diffraction pattern has been obtained by Neider from a few crystals in which reflections corresponding to a large lattice spacing of 22·6 Å appear between the (001) spots corresponding to a 12·12 Å spacing. These spots disappear after a few minutes irradiation by the electron beam leaving only the (001) spots. A corresponding phenomenum appears in the micrograph in that lines with a spacing of 22·4 Å are first seen, but this spacing becomes halved after a few minutes. A similar doubling effect has been observed by Labaw and Wyckoff (1957 a, b) (see below) with indanthrene scarlet, although they do not report any effect of the electron beam on the spacing. No explanation of this effect is available at the present time.

A striking feature of all micrographs of the phthalocyanine class of crystals is the comparative rarity of dislocations. It is natural to assume from this that the density of dislocations is indeed very low. This is probably correct, but it should be remembered that a further consequence of the analysis given above is that no extra half-planes are seen in the image if the Burgers vector is normal to the plane of projection since it then has no component in the plane. This would be the situation if the crystals grow by the Frank spiral mechanism centred on a screw dislocation with Burgers vector normal to the (001) habit surface. No dislocation movement has been observed in the phthalocyanines. This is not surprising in view of the obvious brittleness of larger crystals.

More complex disturbances are sometimes observed in phthalocyanine lattices, but no attempt has yet been made to analyse these in detail. A typical example shown in fig. 21, Pl. 35, was originally interpreted by Menter (1956 a) as possibly a screw dislocation, but the more recent analysis described above shows that no special configuration unique to a screw is expected in the projected image. Infrequently low angle boundaries consisting of arrays of dislocations have been observed in phthalocyanines. Figure 22, Pl. 36, shows a crystal which may be polygonal rather than continuously curved. Any dislocations associated with the polygonal form would have Burgers vectors parallel to the resolved planes and are therefore not detected on the micrograph.

4.3.2. *Indanthrene scarlet R*

Crystals of indanthrene scarlet R have been studied by Labaw and Wyckoff (1957 a, b). The molecule has the composition $C_{38}H_{22}N_2O_6$ and structure

Specimens prepared by direct sublimation on to grids without supporting films consisted of needle-like crystals, many of them only 200–300 Å in thickness. Spacings of 15·4, 19·3 and 28·1 Å have been resolved. Micrographs of the 19·3 Å spacing are very similar to those obtained with the phthalocyanines showing black and white lines of uniform width. The other spacings show some new features. In some crystals with the 15·4 Å spacing alternate dark lines are especially black or alternate white lines especially transparent (fig. 23, Pl. 37). With the 28·1 Å spacing the dark lines are frequently wider than the spaces between them and sometimes the dark lines split completely to give a uniform 14·0 Å spacing (fig. 24, Pl. 38). No x-ray data are at present available for this compound but the authors assume that the three spacings resolved are characteristic of the structure since particularly good internal consistency is obtained in the measurement of any one of them. They have, for example, been able to distinguish the 14·0 Å ' half-spacing ' from the 15·4 Å spacing. It may be that the intensity modulation effects observed are associated with higher order spectra contributing to the image so that it might be informative to study the effect on such images of varying the objective aperture size.

A further new phenomenum observed in this work is shown in fig. 24 (*e*). The abrupt change in direction of the lattice near the end of the crystal is associated with a twin orientation.

This study of indanthrene scarlet R marks a significant point in development of methods of determining crystal structures. It is the first time that the approach to the structure of a crystal with spacings within the range of fairly easy diffraction measurement has been made by the alternative path of direct microscopic observation.

4.3.3. *Molybdenum trioxide*

Bassett and Menter (1957) have investigated the region below 10 Å to determine the smallest lattice that can be directly resolved with currently available instruments. Although the analysis of § 4.2.2 suggests that considering the image impaired only by spherical aberration it should be possible to resolve distances as small as 2–3 Å, this investigation was limited to the 5–10 Å region since it was thought that other adventitious factors might become predominant at the extreme resolution and so mask the true performance of which the instrument is capable. There are not a great number of materials with lattice spacings within the 5–10 Å region which meet all the requirements for lattice resolution, but some success has been obtained with molybdenum trioxide. This crystal has perfect cleavages on (010) and (100) and a less perfect but distinct cleavage on (001). It is possible, albeit infrequently, to obtain tabular crystals which are thin in the [001] or [100] directions. With these it is possible to resolve the (020) spacing of 6·9 Å. Figure 25, Pl. 36, shows a micrograph of a small flake in which this spacing is clearly resolved. This is the smallest

directly resolved spacing yet reported. It should be pointed out, however, that smaller periodicities have been observed in moiré patterns from overlapping crystals (see below).

4.3.4. *Faujasite*

Synthetic faujasite is a sodium aluminosilicate with a composition approximating to $Na_3Al_3Si_4O_{04}$ (8·2–10.1 H_2O). X-ray powder diffraction patterns indicate that the structure is cubic with a diamond type space group and unit cell dimension $a_0 \sim 24{\cdot}8$ Å. The structure is essentially an open aluminosilicate framework (Bergerhoff 1956, Barrer 1956, Barrer *et al.* 1957). In low magnification electron microscope studies Gard (1956) found that larger crystals (1–5 μ) were frequently cubic in habit and opaque to the electron beam. Some smaller particles with less well-developed external form were sufficiently thin for observation of the diffraction pattern and Menter (1957) examined some of these at high resolution. Figure 14 shows a typical micrograph in which the (111) spacing of $\sim 14{\cdot}4$ Å is resolved. Being cubic this structure presents the possibility of simultaneously resolving two sets of lattice planes, by observing for example along [110]. Figure 26, Pl. 39, is a micrograph from a crystal in this orientation. The two sets of (111) planes intersecting at 70° are clearly seen. A model of the structure viewed in this direction shows channels about 8 Å in diameter, 16 Å apart, running right through the crystal. In a limited sense as Gard (1956) suggests, one may regard the image as a direct visualization of the projection of these channels.

Another crystal within this class described by Gard as sodium A has been studied by Bassett (unpublished). X-ray powder photographs by Bultitude (see Gard 1956) indicate a primitive cubic unit cell with $a_0 \sim 12{\cdot}3$ Å. The (100) spacing has been resolved on thin edges of particles.

4.3.5. *Antigorite*

It has been shown above that the recombination of coherent diffracted beams from crystal planes produces an image with a periodicity corresponding to the spacing between the lattice planes causing diffraction. It is to be expected therefore that a superlattice may be resolved if the structure factor associated with the superlattice reflections is sufficiently great to give a reasonable diffracted intensity. Recently Brindley *et al.* (1958) have obtained micrographs from the mineral antigorite (Yu Yen stone) in which a superlattice spacing of ~ 100 Å is resolved (fig. 27 (*a*), Pl. 39). Figure 27 (*b*), Pl. 39, shows an electron diffraction pattern from an antigorite crystal in which the closely spaced satellite reflections about the main lattice reflections are clearly resolved. These satellites also form about the zero order beam and it is these which recombine to produce the periodic image. Since the angles between these beams and the axis are very small the effect of lens aberrations is very slight and it is to be expected that several orders of reflection will recombine to give much sharper lines in the image than is observed with the small lattice spacings

previously discussed. The 40 Å spacing reported by Brindley *et al.* on diffraction patterns from other types of antigorite has also been resolved on micrographs by Chapman† and Menter (unpublished). The authors suggest that the superlattice is a modulation of the scattering factor of the lattice rather than the lattice parameter. Dislocation like features are seen on micrographs of these crystals and there is clearly a field of considerable interest to be developed here on the structure of dislocations in superlattices.

4.4. *Limiting Factors in Direct Resolution of Lattices*

No detailed analysis has yet been made for electron lenses of the combined effect of lens aberration on the image of a periodic object. The elementary treatment of the effect of spherical aberration given above seems to indicate that this at least is not a serious limitation. This is consistent with the analysis of Hopkins (1953) for imaging of periodic objects with light. Residual astigmatism is unlikely seriously to impair the image of a singly periodic object provided the elongation of image points arising from the aberration is in the same direction as the lines of the object. It may be then, considering only the lens aberrations as sources of image deterioration that the chromatic error is the most serious. At the moment there is insufficient experimental evidence concerning the detailed processes of energy loss in the specimen for any reliable, theoretical analysis to be made of the likely effects on the image. Whatever may be the outcome of more detailed theoretical considerations, experimental evidence given above shows that a single periodic structure of spacing 6·9 Å may be resolved with axial illumination. (Later evidence is given for extending this to 5·8 Å.) On the assumption that resolution limitation processes at the moment arise from lens aberrations which are all related to some power of the lens aperture, following the Abbe theory we may expect that the use of oblique illumination should extend the resolution limit by a factor of two to about 3 Å. With the zero order beam and first order on one side only, equally inclined to the axis of the lens the effect of aberrations should be no more serious than in the symmetrical situation with both first order beams from a lattice of spacing 6 Å contributing to the image. Although a 3 Å spacing could not be visually observed on the fluorescent screen, the instrumental magnification and plate resolution should be ample for resolving it in a through-focus series.

Considering effects other than those arising from the lens aberrations the two most serious are specimen movement and contamination. With care, nitrocellulose supporting films can be adequately prestabilized by irradiation in the electron beam, for work down to 5–6 Å although it must be admitted that the poor percentage yield of good pictures is due as much as anything to residual specimen movement, the exact causes of

† Private communication from J. Zussman.

which have not been systematically investigated. Self supporting metal films such as those described later are much better in this respect and it may be advantageous in some circumstances to use these as supporting films for other specimens. Under certain conditions of growth it is possible to obtain these metal films with a high proportion of holes. Crystals deposited on these and suspended across the holes may be examined without the confusion of background structure from the support.

The most serious result of specimen contamination is the rapid loss of contrast in the image. Ennos (1953, 1954) has shown that the main source of contaminant is hydrocarbon matter condensed from the vapour phase on to the specimen, where it suffers degradation into a graphite-like structure under electron bombardment. In typical conditions of observation, the layer may grow at a rate of several Å/sec. Leisegang (1956 b) and Leisegang and Schott (1957) have shown that the contamination may be virtually eliminated by surrounding the specimen with a cooler surface (cooled externally by liquid air) on to which the unwanted vapours preferentially condense. It remains to be seen how well the specimen may be stabilized against thermal drift in these circumstances.

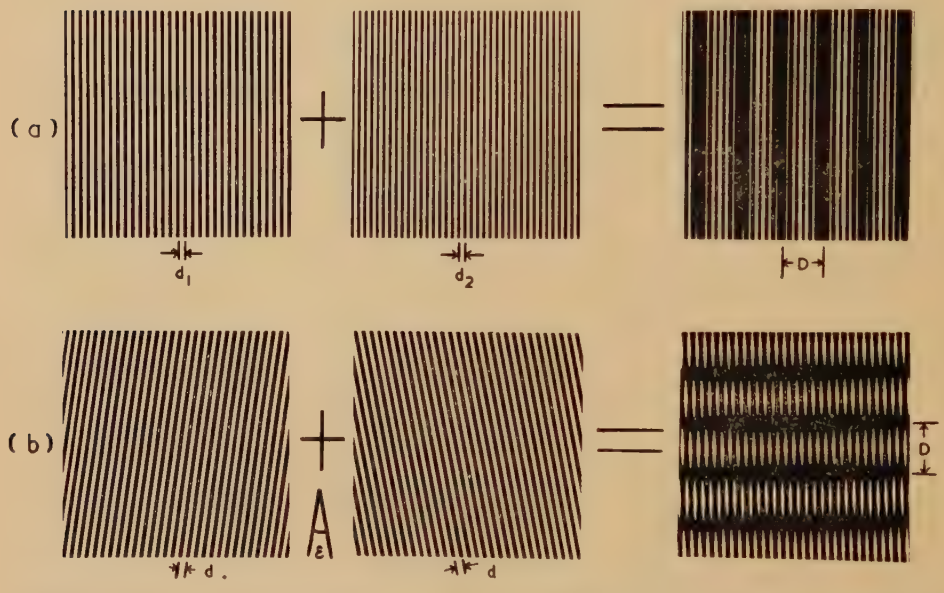
An alternative approach to the imaging of periodic structures has been proposed in a series of papers by Cowley and Moodie (1957 a, b, c). If a periodic object is illuminated by spherical wave fronts from a small source these authors have shown theoretically and by demonstrations with an optical analogue that a series of images are formed in particular planes both in front of and behind the object having the same geometrical form as the periodic object. These images are called Fourier images by analogy with Fourier projections in crystallography. The virtual images on the source side of the object are demagnified but the real images on the other side are magnified. The factor limiting resolution in the light optical case, apart from that set by the wavelength of the radiation used, appears to be the perfection of the object. The proposal is to obtain a few times initial magnification of a crystal lattice using a small electron source and to image the Fourier image thus produced by a conventional high resolution electron microscope arrangement of objective and projector lenses. The main practical problem is to obtain a suitable small electron source. The authors show that it need not be excessively small, although to obtain good contrast this is advantageous, but the edges of the source must be very sharp, with an unevenness considerably less than the least resolvable distance of the best available electron microscopes. Four possible types of source proposed are (1) small electron emitting particles, (2) illuminating a small aperture with an intense electron beam, (3) using electrons scattered elastically from a small, thin particle or from the edges of a thick particle or aperture, and (4) using an electron optical lens to form a demagnified image of a small source or aperture. The experimental difficulties in achieving any of these are clearly formidable.

§ 5. Indirect Resolution of Lattices by means of Moiré Patterns

5.1. *Principles of Moiré Patterns*

It is clear from the foregoing that it may be extremely difficult, with existing instruments, to resolve lattices with dimensions below 2–3 Å. It is fortunate that an indirect method is available, which gives us, at least in the case of metals, almost as much information as would be obtained from the direct method. The technique is to apply the well-known optical principle of forming moiré patterns to the imaging of crystal gratings with electrons. Two basic types of moiré pattern may be formed as illustrated in fig. 28 by the superposition of bar and space gratings.

Fig. 28



Two basic types of moiré pattern. (*a*) Parallel. Moiré spacing $D=d_1d_2/(d_1-d_2)$. (*b*) Rotation. Moiré spacing $D=d/\epsilon$.

The parallel moiré pattern is formed by the parallel superposition of two unequal gratings of pitch d_1 and d_2, giving rise to a moiré pattern of lines parallel to the lines of the elementary gratings with a spacing

$$D=d_1d_2/|d_1-d_2|. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The rotation moiré pattern is formed by the superposition of two equal gratings of pitch d with a small angular twist ϵ about the normal to the plane of the gratings. This leads to a moiré pattern the direction of which bisects the obtuse angle between the lines of the elementary gratings, and

has a spacing given by

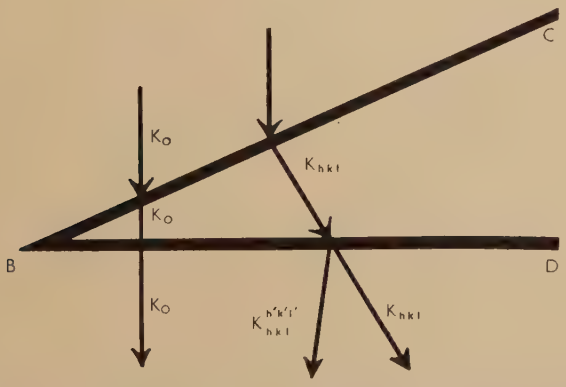
$$D = d/\epsilon \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Combinations of the two types may of course be obtained, together with patterns arising from gratings with angular misorientations other than the simple one described above, but for present purposes we shall confine our attention to the simple cases. For a simple grating of the type used in the illustration, the moiré pattern is an enlarged version of the elementary gratings in one case parallel and in the other approximately perpendicular to them. We may therefore define a ' moiré magnification ' in the parallel case of $d_2/d_1 - d_2$ and in the rotation case of $1/\epsilon$. If a microscopic system is incapable of resolving the elementary distance of a grating, by superposing a second grating of suitable pitch or relative orientation on the first, we can produce a coarser pattern, similar in form to the elementary gratings with a spacing that is within the resolution limit of the microscope. This principal may be applied to the resolution of crystal lattices with the electron microscope.

5.2. *Electron Microscope Images of Overlapping Crystals*

During recent years several workers have reported uniformly spaced fringe systems in electron microscope images of overlapping crystals. They were first noticed by Mitsuishi *et al.* (1951) in specimens of graphite.

Fig. 29



Double diffraction by superposed crystals misoriented by rotation about an axis normal to plane of page. Mitsuishi *et al.* (1951).

These authors based an explanation on the assumption that partial cleavage of a graphite flake could produce a geometry similar to that shown in fig. 29 where the two crystal flakes BC, BD form a small wedge angle. If the incident wave vector is K_0, the vectors of the beams emerging on the far side of the crystal BC are K_0 and K_{hkl}, the latter being a primary diffracted beam from the (hkl) planes of this crystal. These two beams

enter the second crystal which is assumed to be oriented so that the beam K_{hkl} is incident at the Bragg angle on a *different set* of planes $(h'k'l')$ in the crystal BD and gives rise to a secondary or doubly diffracted beam with vector $K_{hkl}^{h'k'l'}$ so that there are three beams K_0, $K_{hkl}^{h'k'l'}$ and K_{hkl} emerging from the second crystal. If the indices hkl, $h'k'l'$ are suitable then the angle between K_0 and $K_{hkl}^{h'k'l'}$ may be very small so that these two beams overlap in the space immediately behind the crystals, and since they have both been elastically scattered they are coherent and lead to the formation of an interference pattern.

The angle α between K_0 and $K_{hkl}^{h'k'l'}$ is given by

$$\alpha = 2|\theta_{hkl} - \theta_{h'k'l'}| = \lambda \left\{ \frac{1}{d_{hkl}} - \frac{1}{d_{h'k'l'}} \right\} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where θ_{hkl}, $\theta_{h'k'l'}$ are the Bragg angles corresponding to the lattice spacings d_{hkl}, $d_{h'k'l'}$.

The spacing S of the fringe system formed by the overlapping beams is then given by

$$S = \frac{\lambda}{\alpha} = \frac{d_{hkl} d_{h'k'l'}}{d_{hkl} - d_{h'k'l'}} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Taking (hkl) as (200) and $(h'k'l')$ as (201) they find for graphite $S = 110$ Å in close agreement with the experimentally determined value of the fringe spacing. The essential diffraction origin of the pattern was confirmed by its occurrence only in those regions of the specimen showing diffraction contrast.

Gard (1956) has observed similar patterns in overlapping crystals of muscovite, tobermorite and potash felspar (see also Gard *et al.* 1955, Gard and Taylor 1957). By taking stereophotographs he was able to confirm that one crystal was so inclined as to make a wedge angle with the other as proposed by Mitsuishi *et al.* and he advanced an explanation along similar lines. He obtained spacings between 100 and 250 Å and confirmed that these were independent of accelerating voltage (i.e. λ) and that their formation was sensitive to the inclination of the pair of crystals to the electron beam, appearing only in regions of the specimen showing diffraction contrast.

Seki (1951, 1953) observed fringes in overlapping crystals of the mineral sericite, with spacings of 360 Å and 230 Å respectively. He attributed them to interference between the undeviated beam and a beam doubly diffracted by the two crystals in succession as a result of a small *rotational* misorientation ϵ of one crystal with respect to the other, about an axis normal to the habit plane of the crystals. The doubly diffracted beam emerges at a small angle to the undeviated beam and the spacing of the interference fringes formed in the region where the beams overlap is given simply by $S = d/\epsilon$ where d is the spacing of the reflecting planes, now assumed to belong to the same form in each crystal (hkl, and $\bar{h}\bar{k}\bar{l}$). Measurement of S and ϵ from the micrographs gave good agreement

between the value of d deduced from the relation above and the value determined from an electron diffraction pattern of a single crystal. He also showed that the fringes are perpendicular to the net planes of spacing d.

Parallel fringes were observed by Bernard and Pernoux (1953) spaced 500 Å and 200 Å apart in images of overlapping crystals of lead oxide and molybdenum oxide respectively. They postulate a configuration of the crystals similar to that proposed by Mitsuishi *et al.* for graphite. Their proposed mechanism implies that the doubly diffracted beam emerging from the second crystal is parallel with the primary beam and that the interference fringes arise from the continuously varying path difference in the wedge-shaped gap between the two crystals in a manner analogous to the formation of wedge fringes in optical systems. These authors had no independent means of confirming exactly the expression deduced for the spacing on this assumption and Dowell *et al.* (1956) have pointed out that this expression is inconsistent with that deduced by them from a more general reciprocal lattice treatment of the double diffraction.

Hillier (1954) observed similar phenomena in platelike crystals of iron oxide but suggested that there was an additional effect which could not be explained on the basis of double diffraction. This consisted of closely spaced fringes extending over very large areas of crystals. There were usually two superimposed sets of fringes with equal spacings crossing at 120°, the spacing being generally below 100 Å with a minimum value of 40 Å. Hillier could not justify a double diffraction explanation requiring rigid satisfaction of the Bragg diffraction condition over such large areas (several microns square). As an alternative explanation he suggested that this was a moiré pattern produced by the superposition of the two crystal gratings with a small relative rotation about an axis normal to the plane of the specimen. (See also Green and Weigle 1948.) The effect was pictured as analogous to the light optical effect obtained by Rayleigh (1874 a, b) by superimposing two coarse bar and space gratings with a slight relative rotation (fig. 28 (*b*)).

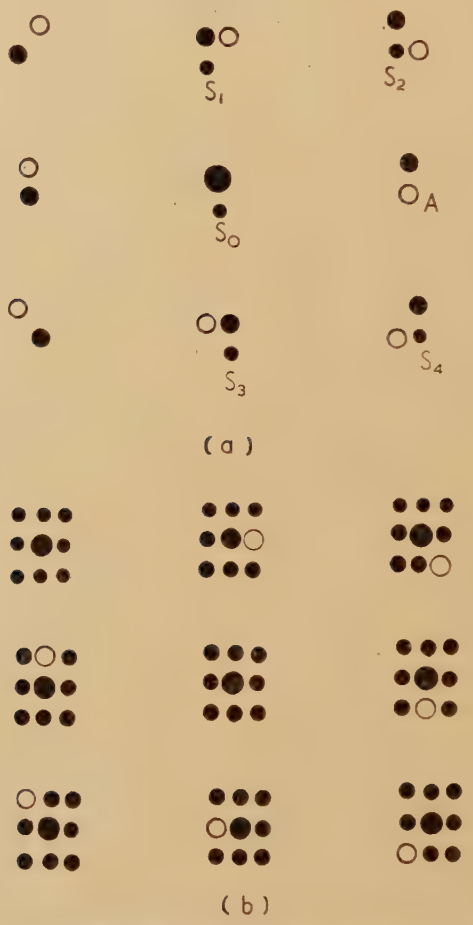
In fact, double diffraction and the simple moiré phenomena from superimposed bar and space gratings cannot be separated in this way. Guild (1956) in considering moiré phenomena in crossed optical diffraction gratings has pointed out that moiré patterns may be formed by pure *phase gratings* in which there is no question of amplitude reduction and that all moiré phenomena must be considered essentially as arising from double diffraction by the two gratings in series. The Rayleigh system then appears as a limiting case which may be treated by simple geometrical superposition only because of the large spacing of the gratings.

All of the fringe phenomena reported by the workers mentioned above are undoubtedly moiré effects, and it had been realized by some that an explanation of the mechanism of fringe formation lay in an understanding of double diffraction. A considerable clarification came with the treatment of Dowell *et al.* given below which brings out clearly the intimate relationship between double diffraction and the formation of moiré patterns.

5.2.1. *Double diffraction and image formation with overlapping crystals*

Since a relatively high proportion of the incident energy appears in the diffracted beams when an electron beam is elastically scattered by a crystal it is to be expected that secondary or double diffraction may occur if a second crystal is placed in the path of the primary diffraction pattern from the first crystal. Effects attributable to this process have been

Fig. 30



(*a*) Origin of doubly diffracted beams from two overlapping crystals with small rotational misorientation about axis normal to page. (*b*) Complete double diffraction pattern built up by considering in turn each diffracted beam from crystal 1 as new primary beam for diffraction by crystal 2. Crystal 1 primaries ○, crystal 2 primaries ●, doubly diffracted beams ●.

recognized since the earliest days of electron diffraction (see, for example, Raether (1932), Taylor-Jones (1934), Finch and Wilman (1937), Darbyshire and Cooper (1935), Finch and Sun (1936), Wilman (1940), Pashley (1951), Cowley *et al.* (1951).

Figure 30 shows the diffraction patterns from two parallel crystals of orthorhombic symmetry (unit vectors **a**, **b** normal to the incident beam) rotationally misoriented by an angle ϵ about the direction of the incident beam. If crystal 1 is nearer the electron source then a primary beam such as A from this crystal passes into crystal 2 where, if the diffraction conditions are sufficiently relaxed, it will act as a new ' zero order ' beam for diffraction by crystal 2 and produce a series of secondary beams S_0, S_1, S_2 . . . . by diffraction from the net planes of crystal 2. By considering in turn each of the primaries from crystal 1 as sources for diffraction by the net planes of crystal 2 we can build up around each of the primaries a miniature cross-grating secondary diffraction pattern (fig. 30 (*b*)). From the point of view of image formation in the electron microscope, special interest attaches to the secondary beams around the original zero order beam since they are inclined at a very small angle to the axis of the instrument if ϵ is small, and may therefore be expected to be coherent with the zero order beam and able to recombine in the image space to give a recognizable periodic pattern. The intensity distribution in this pattern may be determined using a treatment given by Dowell *et al.* (1956). The two dimensional potential distribution of one lattice may be represented by

$$\phi(x,y)=\sum_h\sum_k V_{hk0}\exp\left\{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)\right\} \quad . \quad . \quad . \quad . \quad (11)$$

Let $\phi(x',y')$ and $\phi(x'',y'')$ denote the respective potential distributions of the two crystals and transform their coordinates to axes x,y, forming the bisectrices of the angle between x' and x'' and y' and y'' respectively. Then the function $\phi(x',y')\,.\,\phi(x'',y'')$ represents the intensity distribution arising from double diffraction. The only spectra of practical interest are those around the zero order since the remainder are excluded by the physical aperture in the back focal plane of the lens. It is readily shown that these correspond to double diffractions for which $h_ik_i=\bar{h}_j\bar{k}_j$ and for a centro-symmetric system ($V_{hk}=V_{\bar{h}\bar{k}}$), the intensity distribution in the image has the form

$$\sum_h\sum_k V_{hk}{}^2\cos 2\pi\epsilon\left(\frac{kx}{b}-\frac{hy}{a}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

This is a periodic function with unit translation $\mathbf{a}/\epsilon$, $\mathbf{b}/\epsilon$ rotated through 90° with respect to the bisectrix of the angle between the *hk*o planes of the two crystals. It is in fact the moiré pattern given by the simple Rayleigh theory for the bar and space grating. The moiré pattern can thus be regarded as an Abbe image formed by the recombination of diffracted beams in a manner analogous to the formation of an image of an elementary grating as described in § 4.2.2. The moiré pattern arises by the recombination of beams with a ' Bragg angle ' corresponding to that expected from a pseudo-grating with the spacing of the moiré pattern. It is important to note that the reciprocal lattice treatment demands no assumption that the primary beam from the first crystal shall be incident on the second at the Bragg angle precisely as had been supposed by Mitsuishi *et al.* As long as the extension of the reciprocal

points due to the thinness of the crystals is sufficient to ensure that they are cut by the sphere of reflection, then double diffraction can occur. Thus, misorientations about other axes will give rise to double diffraction and Dowell *et al.* give the following expressions for two further simple cases.

(ii) For a tilt about an axis normal to the reciprocal lattice vector forming a wedge angle α between the crystals, the spacing of the moiré pattern is given by :

$$S=\lambda/\alpha^2\theta=2d/\alpha^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

and the fringes are perpendicular to the diffracting plane normal.

(iii) For a tilt α_n about the diffracting plane vector and a deviation β from the true Bragg setting

$$S=\lambda/2\beta\alpha_n\theta=d/\beta\alpha_n. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

The fringes in this case lie parallel to the diffracting plane normal.

In general for specimens consisting of plate-like crystals deposited at random on a supporting film the situation with a small rotational misorientation about the normal to the habit planes will be most common. The case of parallel unequal lattices may be treated in a similar manner (Pashley, unpublished). Figure 31 (*a*) shows the secondary beams S_0, S_1, S_2 . . . . produced by one of the primary beams P from crystal 1 acting as a new source for diffraction by crystal 2 and fig. 31 (*b*) shows the total diffraction pattern obtained by considering all possible double diffractions. As before, a miniature cross-grating diffraction pattern is formed around the zero order beam leading to an intensity distribution in the image with a form given by

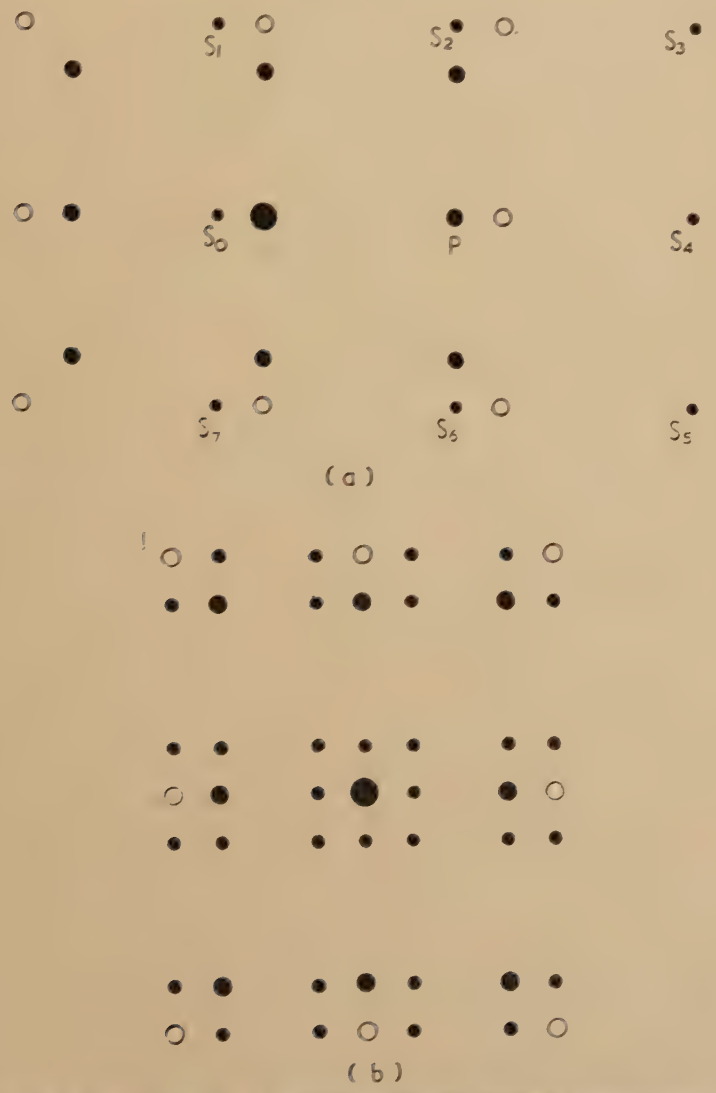
$$\sum_h\sum_k V_{hk}{}^2 \exp\left\{2\pi i\left(\frac{a_1-a_2}{a_1a_2}hx+\frac{b_1-b_2}{b_1b_2}ky\right)\right\} \quad . \quad . \quad . \quad (15)$$

where $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{b}_1$, $\mathbf{b}_2$ are the unit vectors of the two lattices in the plane of projection. This represents a moiré pattern with unit translations $\mathbf{a}_1\mathbf{a}_2/\mathbf{a}_1-\mathbf{a}_2$, and $\mathbf{b}_1\mathbf{b}_2/\mathbf{b}_1-\mathbf{b}_2$ in directions parallel to a_1, a_2 and b_1, b_2 respectively.

The essential principles of indirect lattice resolution may be readily demonstrated using an optical analogue. Using an arrangement similar to that shown in fig. 1 diffraction patterns are recorded in the back focal plane of the lens imaging the overlapping gratings on a screen in the image plane. Dot gratings of pitch 220 and 180/inch respectively were used in these experiments. Figures 32 (*a*, *b*), Pl. 40, shows one grating with its diffraction pattern and fig. 33, Pl. 40, shows the double diffraction pattern from the two gratings superposed with their elementary vectors parallel. The double diffraction spots around each primary are clearly seen. Figure 34 (*a*), Pl. 41, is the image of the superposed gratings with no stop in the back focal plane of the lens so that all the diffracted beams contribute to the image. Figure 34 (*b*), Pl. 41, shows the effect of placing a stop in the back focal plane to exclude all beams except the zero order and the secondaries immediately surrounding it. The image is now a moiré pattern with a pitch of 40 inch. The stopped down lens simulates the situation in the electron microscope. With a full aperture and

aberration-free lens we would obtain an image analogous to fig. 34 (*a*), in which all the fine detail of both gratings is clearly resolved. With a limited aperture, as must necessarily be used because of the imperfection of the electron lens, the fine detail is removed from the image but a simple pattern identical in form to the originals is obtained. A similar demonstration may be made for rotated gratings. Figures 35 (*a*, *b*, *c*), Pl. 41, show respectively the diffraction pattern, image with full aperture and image

Fig. 31



(*a*) Origin of doubly diffracted beams from two parallel overlapping crystals with unequal lattice parameters. (*b*) Complete double diffraction pattern built up by considering in turn each diffracted beam from crystal 1 as new primary beam for diffraction by crystal 2. Crystal 1 primaries ●, crystal 2 primaries ○, doubly diffracted beams ●.

with limited aperture of two superposed gratings (200/inch) with a relative rotation of 10°. A direct demonstration of the effect of aperture on an electron micrograph where both the elementary lattice and the moiré pattern may be resolved can be made with the phthalocyanines. Figure 53 (*a*), Pl. 53, shows two overlapping crystals of platinum phthalocyanine in which the $(20\bar{1})$ spacing and the corresponding moiré pattern are resolved. In fig. 53 (*b*), Pl. 53, a small objective aperture has been inserted to exclude the $(20\bar{1})$ spectra so that the moiré pattern alone is resolved.

5.2.2. *Moiré patterns from molybdenum oxide crystals*

Dowell *et al.* (1956) confirmed their double diffraction interpretation of moiré patterns from observations on fortuitously overlapping crystals of molybdenum trioxide. Some laminar crystals were found to lie one on another with an angle of rotation about the normal to the main habit surface close to 90° (fig. 36 (*a*), Pl. 42). The *a*- and *c*-axes lying in the plane of the crystals are almost equal and double diffraction may arise either from (100) in the first crystal and (001) in the second or from (101) in the first and (101) in the second. The situation is therefore mixed between the simple rotation and parallel cases considered above and a modified expression was deduced for the spacing s.

$$s = \frac{d_1 d_2}{\{d_1^2 \alpha_c^2 + (d_2 - d_1)^2\}^{1/2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where d_1, d_2 are the spacings of the net planes in the first and second crystal respectively and α_c is the angle between the diffracting plane normals.

Measurement of s and α_c from the micrographs gives remarkable agreement with the calculated values of s.

Table 1

	S obs	S calc
$101/10\bar{1}$	105 Å	101 Å
100/001	52 Å	52 Å

The doubly diffracted beams essential for the formation of the moiré pattern were later observed (Dowell *et al.* 1957) very close to the central beam (fig. 36 (*b*), Pl. 42. These could be indexed on the assumption of two crystals rotated 90° with respect to each other.

A direct confirmation of the essential role of double diffraction in forming a bright field moiré pattern has been obtained by Bassett, Menter and Pashley (unpublished) in an experiment with superposed parallel crystals of copper and gold prepared by the method described in § 5.3. Figure 37 (*a*), Pl. 43, shows the moiré pattern obtained using a 50 μ objective aperture which allows the doubly diffracted beams around the zero order to reach the image. In fig. 37 (*b*), Pl. 43, showing the same

field of view imaged with a 5μ aperture to exclude all but the zero order from the image, the moiré pattern has disappeared.

5.2.3. *Physical significance of moiré patterns*

It is important to consider the physical significance of the moiré pattern. Farrant and Rees (1956) have pointed out that the intensity expression for the image of two overlapping gratings with rotational misorientation ϵ is in fact a Patterson distribution rotated through 90° with respect to the function representing a single lattice and having a magnification of $1/\epsilon$. The Patterson function is a vector map showing maxima corresponding to the vectors between the atoms in the crystal, the strength of the maxima being proportional to the product of the scattering powers of the two atoms concerned. It is commonly used in the early stages of a structural determination to locate the positions of the main scattering centres in the unit cell. The interpretation of a Patterson map is easiest in fairly simple structures or structures where there are a few relatively heavy atoms. Since the moiré pattern represents the Patterson function, it is possible by choosing ϵ small enough to image the function at suitable magnification $(1/\epsilon)$ to obtain any predetermined amount of detail. Thus in a microscope which will resolve periodic structures of spacing 10 Å, vector displacements of 1 Å are readily resolved using a value of $\epsilon = 0{\cdot}1$ rad. The correspondence of a moiré pattern with the Patterson distribution has been demonstrated by Farrant and Rees (1956) using an optical analogue. Figure 38 (*b*), Pl. 44, shows the moiré pattern formed by the superposition of two gratings of the form shown in fig. 38 (*a*), Pl. 44. The gross structure of the former is readily identified with the Patterson distribution of the latter. As these authors point out there are experimental difficulties in the way of using this method of recording Patterson functions directly since it is necessary to control the relative orientation of the crystals accurately and to maintain their relative parallelism in the plane of the specimen over reasonably large areas in order to obtain a true rotation moiré pattern. A further limitation to be remembered is the effective termination of the Patterson series for higher orders due to the spherical aberration of the objective lens. In general and particularly with small values of ϵ, this will not be too serious since the diffracted beams pass through the lens at relatively small angles to the axis. We may expect increasing use of this method in the structure analysis of crystals by electron diffraction.

5.2.4. *Lattice resolution by moiré patterns*

With simple lattices with single atoms at each lattice point, the positions of the Patterson peaks are coincident with the positions of the atoms in the unit cell so that the Patterson projection is identical geometrically with the projection of the real lattice. The moiré pattern from cubic and h.c.p. metals is equivalent geometrically to a magnified projection of the atomic lattice itself. In the rotation pattern this projection is rotated 90° with respect to the real lattice projection for identical metals

or some other determinable angle for different metals. In the parallel pattern the projection is parallel to the real lattice projection. The moiré magnification effect for avoiding the limitations of direct lattice resolution described in § 4.4 is thus a powerful means of investigating the atomic structure of metals. By choosing ϵ or d_1 and d_2 appropriately we can always obtain a periodic image within the resolution limit of any instrument. At first sight it might seem advisable always to use as high a moiré magnification as possible. It will be shown later that the magnification should be kept as small as is consistent with the resolution achievable in order to extract the maximum amount of information from the image.

It is of interest to note that moiré patterns provide a ready means of determining the resolution limit of an instrument, at least for periodic objects, since we may design a specimen to produce any predetermined spacing from the net plane spacing upwards. The smallest resolved spacing so far reported is 5·8 Å (fig. 39, Pl. 45) from superposed parallel single crystals of nickel and gold (see below).

5.3. *Moiré Patterns from Thin Metal Films*

All of the earlier observations on moiré patterns reported above has been confined to overlapping crystals formed by chance in conventional specimen preparations. Using the evaporation technique for the preparation of thin single crystal metal films (Pashley 1958), Pashley *et al.* (1957), Bassett *et al.* (1958) have developed controlled methods of producing moiré patterns and have used these to study imperfections in the films.

Thin single crystals of gold in (111) orientation were prepared by epitaxial growth on silver substrates from which they could subsequently be detached by solution of the silver in nitric acid. Moiré patterns were then formed in one of two ways. (1) A second metal film of a different metal was grown epitaxially upon the first so that the unit cell lattice vectors were parallel. (2) A second film, either of the same or different metal, was placed in contact with the first with a small angular twist between the two. The combinations of metals shown in table 2 have been investigated in parallel orientation. The geometry of the moiré patterns obtained can be understood most readily in terms of the diffraction patterns. Figure 40 (*a*), Pl. 44 shows the pattern obtained from a single crystal gold film about 300 Å in thickness. In the (111) orientation the relevant diffracted beams are {220} and {422}. Figure 40 (*b*), Pl. 44 shows the double diffraction pattern produced by these diffracted beams from a specimen consisting of a nickel crystal grown epitaxially upon the gold crystal. The doubly diffracted beams from combinations of $220/\bar{2}20$ and $422/\bar{4}\bar{2}2$ type reflections are clearly resolved. In the moiré pattern we may thus expect to see effects arising from the combination of any of the central secondary beams with the zero order beam. The spacings of the expected moiré patterns calculated from the normal lattice constants are given in table 2. Figure 41, Pl. 46 shows a typical micrograph obtained from a Au/Pd combination. Two main spacings are observed corresponding to double diffraction from (220) and (422) primaries respectively. Three

orientations 120° apart are possible for each spacing, each (422) type being perpendicular to a particular (220) type. The type of moiré pattern seen in a given region of the specimen is determined by the reflecting conditions there. If the specimen is not flat these will vary from one region to another and only in exceptional circumstances will the orientation be suitable to form the moiré pattern from several sets of planes simultaneously to give a characteristic cross hatched or dot pattern. Figure 42, Pl. 47 shows an example of the latter in a Cu/Au specimen. In part of the field shown it appears that at least two and possibly three (220) type reflections are contributing to the image.

The measured spacings of the patterns agrees to within about 10% with the values calculated from eqn. (7) and those deduced from the positions of the doubly diffracted beams. In some cases, however, there are definite discrepancies outside the limits of experimental error. These are probably associated with specific properties of the bimetal film arising from effects particular to the growth of one film on the other, which will be more fully discussed below.

Table 2

Metal	Crystal spacing (Å)		Calculated moiré spacings (Å) when combined with parallel gold	
	$02\bar{2}$	$42\bar{2}$	$(02\bar{2})$	$(42\bar{2})$
Nickel	1·24	0·719	9·2	5·3
Cobalt	1·26	0·725	9·8	5·7
Copper	1·28	0·737	11·3	6·5
Palladium	1·37	0·792	29	17
Platinum	1·38	0·798	35	20

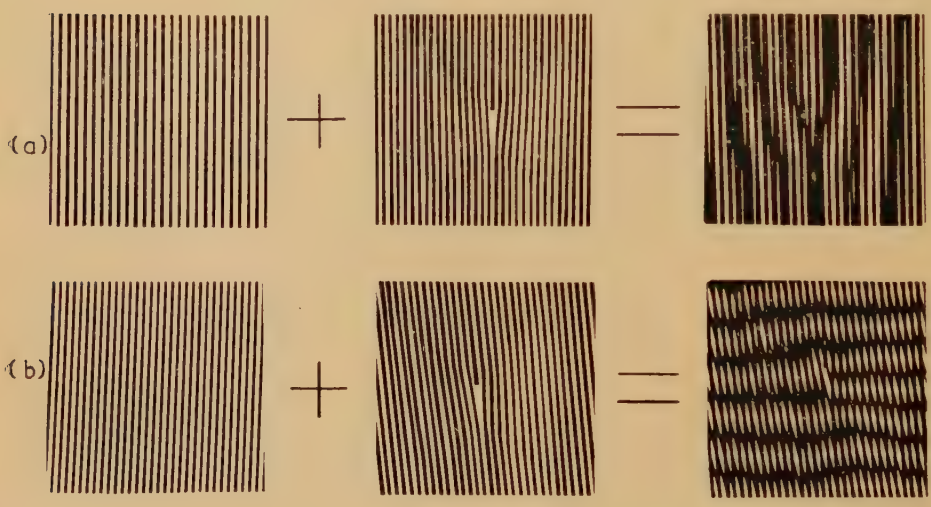
The second method for producing moiré patterns is to be preferred for studying the structure of metal films since the patterns are not likely to be modified by interfacial effects. Figure 43, Pl. 47 shows a moiré pattern from two separately prepared gold films placed one on the other with a relative rotation of $\sim 2°$ about the normal to the plane of the films. Provided good contact can be established between the two films, relatively large areas of good moiré pattern can be obtained as shown in fig. 44, Pl. 48 which is a moiré pattern from superposed separately prepared films of gold and palladium. The double diffraction pattern from a pair of rotated gold crystals (fig. 45, Pl. 44) shows the secondary beams around the central beam responsible for the moiré pattern.

5.3.1. *Dislocations in metals resolved by moiré patterns*

Hashimoto and Uyeda (1957) and Pashley *et al.* (1957) independently realized that dislocations in one or other of the crystals producing a moiré pattern should appear as extra half lines in the pattern. This is readily demonstrated by the optical analogues shown in fig. 46. In the rotation

case the extra half line in the pattern is perpendicular to that in the dislocated grating. In the parallel case the extra half line is parallel to that in the grating, having the same 'sign' if the dislocation is in the coarser grating and opposite sign if it is in the finer grating. As was mentioned in § 4.2.4, in the direct projected image of a lattice the number of extra half lines seen is determined by the resolved component of the Burgers vector of the dislocation line in the plane of projection, perpendicular to the imaged planes. This projection is the effective grating from the point of view of the moiré analogue so that we expect to see the same number of extra half lines in the moiré pattern of a dislocated crystal as would be seen, were it possible to form a direct projected image of the dislocated lattice. The number of these extra half lines for a dislocation with Burgers vector $\frac{1}{2}\mathbf{a}$ [110] in images of face-centred cubic crystals with various planes resolved is shown in table 3.

Fig. 46



From *Nature, Lond.*

Optical analogue showing appearance of dislocations in moiré patterns. (*a*) Parallel case. Pashley *et al.* (1957). (*b*) Rotation case. Hashimoto and Uyeda (1957).

All the dislocation features observed by Pashley *et al.* are consistent with this analysis. Figure 47, Pl. 49 shows a dislocation $\frac{1}{2}\mathbf{a}$ [$1\bar{1}0$] in a moiré pattern from Au/Pd in which the number of extra half lines are 2, 1, 1, 3, 3, 0 on (220), (20$\bar{2}$), (0$\bar{2}$2), (422), (242) and (22$\bar{4}$) respectively. If one set of planes only is resolved it is possible to define the Burgers vector unambiguously only when it is perpendicular to these planes and in the plane of the specimen leading to two extra half lines (see, for example, figs. 41, 47). In this particular orientation of the specimen all dislocations will be detected provided at least two sets of planes are resolved so that reasonably reliable estimations of dislocation density may be made for

separately prepared films. An approximate figure of between 10^{10} and 10^{11} cm^2 has been determined for these vacuum evaporated films. It is, of course, not possible by inspection to determine in which crystal the dislocation lies. Moreover it is a consequence of the sign reversal effect described above that dislocations of the same type immediately one above the other will cause no extra half lines to appear in the moiré pattern. This situation may arise quite frequently when one crystal is grown upon the other, since a dislocation present in the first crystal may be continued in the second. Figure 48, Pl. 50 illustrates a situation where this may have occurred. In fig. 48 (*a*) no extra half lines are seen in the enclosed

Table 3

Plane (*hkl*)	(220)	(202)	(022)	(2$\bar{2}$0)	(20$\bar{2}$)	(02$\bar{2}$)
Number of extra half-lines (*N*)	2	1	1	0	1	1
Plane (*hkl*)	(422)	(242)	(224)	(42$\bar{2}$)	(4$\bar{2}$2)	($\bar{4}$22)
Number of extra half-lines (*N*)	3	3	2	3	1	1
Plane (*hkl*)	(24$\bar{2}$)	(2$\bar{4}$2)	($\bar{2}$42)	(22$\bar{4}$)	(2$\bar{2}$4)	($\bar{2}$24)
Number of extra half-lines (*N*)	3	1	1	2	0	0

region but in fig. 48 (*b*) taken a few seconds later after some dislocation movement had been induced (see below) two extra half lines of opposite sign are seen. This may be interpreted by supposing that a dislocation in the first crystal was copied during growth of the second and that the subsequent stress has separated the line into two segments, one in the upper crystal and the other in the lower, one dislocation now appearing in the moiré pattern with its true sign and the other with reversed sign as expected.

5.3.2. *Partial dislocations and stacking faults*

In the face-centred cubic system a perfect dislocation $\frac{1}{2}\mathbf{a}$ [110] may dissociate into two imperfect or partial dislocations producing a plane of stacking fault bounded by the partials, called an extended dislocation, according to the reaction

$$\tfrac{1}{2}\mathbf{a}[110] \rightarrow \tfrac{1}{6}\mathbf{a}[12\bar{1}] + \tfrac{1}{6}\mathbf{a}[211].$$

By applying to the partials the same analysis as was used to determine the number of extra half lines in a projected image expected from a perfect

dislocation of given Burgers vector it is possible (Bassett *et al.*) to deduce the expected appearance of extended dislocations. The results for the above reaction are given in table 4. It will be seen that some orientations give rise to fractional values for $N(=hu+kv+lw)$, although the sum for the two partials is of course always integral as was found in the previous analysis of perfect dislocations. This means only that the planes are fractionally displaced at either end of the fault so that a real physical significance can be attached to these values. An example of the effect is shown in fig. 49, Pl. 50 from two nearly parallel crystals of gold and palladium in which a displacement of $+\frac{1}{3}$ is observed at one end of the stacking fault and $+\frac{2}{3}$ at the other end. Thus in the orientations of

Table 4

P $[\frac{1}{2}\,\frac{1}{2}\,0]$		P_1 $\frac{1}{6}[12\bar{1}]$	P_2 $\frac{1}{6}[2\bar{1}1]$
(220)	2	1	1
(202)	1	0	1
(022)	1	$\frac{1}{3}$	$\frac{2}{3}$
(2$\bar{2}$2)	0	$-\frac{1}{3}$	$+\frac{1}{3}$
(20$\bar{2}$)	1	$\frac{1}{3}$	$\frac{2}{3}$
(02$\bar{2}$)	1	$\frac{2}{3}$	$\frac{1}{3}$
(422)	3	$\frac{4}{3}$	$\frac{5}{3}$
(224)	2	1	1
(242)	3	$\frac{5}{3}$	$\frac{4}{3}$
(42$\bar{2}$)	3	$\frac{4}{3}$	$\frac{5}{3}$
(4$\bar{2}$2)	1	0	1
($\bar{4}$22)	−1	0	−1
(2$\bar{2}$4)	0	$-\frac{1}{3}$	$+\frac{1}{3}$
($\bar{2}$24)	0	$+\frac{1}{3}$	$-\frac{1}{3}$
(22$\bar{4}$)	2	1	1
(2$\bar{4}$2)	−1	−1	0
($\bar{2}$42)	1	+1	0

table 4 where both partials are resolved it is possible to make a direct measurement of the width of dislocations. The measurement can be made only to within one moiré spacing however, and this is one reason why one should always use as low a moiré magnification as possible. For example in fig. 41 the encircled dislocation has its two extra half lines immediately adjacent in the moiré pattern. Since the moiré spacing is ~ 30 A the detail of the atomic configuration around the dislocation is indicated only on a fairly coarse scale. With a smaller spacing in the pattern one might well find that the two extra half lines marking the partials were separated by a definite number of continuous lines so that the width of the dislocation is then more accurately defined. An example of the latter effect is shown in a parallel Cu/Au combination in fig. 50, Pl. 51 where the two extra half lines are separated by three continuous lines, permitting the specification of the dislocation width as ~ 40 Å. As noted before, this may refer either to the copper or the gold as we do not know

in which film the dislocation lies. In some orientations, as can be seen from table 4, the stacking fault appears only as a result of the deflection of the moiré pattern at either end, if the net effect of the perfect dislocation produces no extra half lines in the pattern. This situation is not so readily characterized on the micrographs although suggestive evidence can be found, and this may well be the explanation of apparent deflections in lattice planes observed in some directly resolved lattices (e.g. faujasite, fig. 14).

5.3.3. *Diffraction contrast of dislocations and correlation with dislocations in moiré patterns*

Hirsch *et al.* (1956) showed that dark lines observed in transmission electron micrographs of thin foils of aluminium passing from one surface to the other or forming more complex networks inside the foil arise from dislocation lines in the metal. In favourable orientations the elastic strain field surrounding a dislocation line causes electrons to be scattered outside the objective aperture so that a cylindrical volume of metal around the dislocation appears 'opaque' in the image giving rise to a black line on the printed image. Figure 51, Pl. 51 shows a number of isolated dislocation lines revealed in this way in a thin film of platinum (Menter, unpublished). If the dislocation line is inclined steeply with respect to the specimen surface then the contrast associated with it degenerates into a diffuse dot. This is commonly the case in thin (111) metal films formed by vacuum evaporation where the slip planes are inclined at 70° to the specimen surface. Figure 52, Pl. 52 shows a typical (111) film of gold about 300 Å in thickness (Pashley 1958) with a number of dislocations revealed in this way. This method of observing dislocations is essentially complementary to the moiré technique described above. It is to be expected that it should be possible to observe the diffraction contrast effect of Hirsch *et al.* superimposed upon a dislocation in the moiré pattern. A number of examples of the association of the two are shown in fig. 43, Pl. 47 which is a micrograph of two separately prepared gold crystals giving a rotation moiré pattern. The coincidence of extra half lines in the moiré pattern with diffraction contrast is by no means universal (see, for example, fig. 41). The reasons for this are not understood and it would be useful for a more detailed investigation to be made of the conditions under which correlation may be expected.

5.3.4. *Observations on dislocation motion*

Hirsch *et al.* observed that dislocation movement could be induced by focusing the illumination into a small area on the specimen and using this technique have made many interesting observations on movement and interaction in aluminium (*loc. cit.*) and stainless steel (Whelan *et al.* 1957). Movement induced in this way may also be observed in moiré patterns (Bassett *et al.* 1958) but for the interpretation of phenomena involving extensive arrays of dislocations the former method is to be preferred since observation can be made at relatively low magnification on the screen

($\sim \times 20\,000$). This has the advantages of a relatively large field of view and adequate brightness for cine recording of the image on the fluorescent screen. In order to observe dislocation motion in moiré patterns a very high magnification ($\sim \times 500\,000$) must be used with a consequent reduction in image brightness and field of view. There is undoubtedly valuable information to be obtained about the detail of dislocation motion at the level of resolution provided by the moiré patterns, but it appears at the moment rather difficult to record this in a manner suitable for subsequent analysis.

5.3.5. *Effect of interface on moiré patterns from metal films*

The analysis given in §5.2.1 of the origin of moiré patterns is based upon the assumption that diffraction by the two lattices is unaffected by the properties of the interface between them. On this view the image is considered as a projection on the plane of the specimen of the projections of the diffracting lattice planes. This is probably true for the situation in which crystals are prepared separately and then placed one upon the other. When the second crystal is grown deliberately upon the first, however, there are a number of factors associated with the growth process itself which may complicate the simple picture.

With metal films alloying may occur at the interface and this may be extensive if the specimen is maintained at an elevated temperature for an appreciable time. Interdiffusion of the two layers causes the lattice parameters to move towards each other resulting in an increased moiré spacing. The situation is by no means simple, however. Unpublished evidence has been obtained in the author's laboratory that in some circumstances triple layers may be formed consisting of two outer layers of composition approximating the pure metals and an inner layer with a well defined alloy composition, leading to two types of superposed moiré patterns, one from the combined effects of the two outer layers and the other from the inner and lower layers. This situation can usually be understood at least qualitatively by studying the diffraction pattern. In other circumstances, however, one may expect a gradual change in lattice parameter normal to the interface. It is not clear at the moment what effect this may be expected to have upon the moiré pattern.

5.4. *Moiré Patterns from Internal Misorientations in Crystals*

All the moiré effects so far described have been obtained from two overlapping crystals. Goodman (1957) has obtained evidence that a moiré pattern may also be formed as a result of internal stacking disorder in a crystal such that one part of the crystal is rotationally misoriented with respect to the other. In studying crystals of the hexagonal modification of boron nitride, Goodman found fringe patterns which could be interpreted in this way. Dark field observation showed that any particular component of the pattern is normal to the doubly diffracting lattice planes in the crystal as would be expected from a true rotation moiré pattern. X-ray studies by Pease (1952) showed that boron nitride

has a layer structure, with hexagonal sheets of B–N layers stacked one above the other along the hexagonal axis. The x-ray pattern indicated that the structure is heavily faulted on the (000*l*) planes and the moiré patterns obtained by Goodman suggest that this faulting consists of small rotations of lamellar sheets about the hexagonal axis.

Similar effects have recently been observed by Keller and O'Connor (1958) in crystals of low molecular weight polythene growing with a platy habit. Successive layers of the crystal are rotationally misoriented with respect to each other giving rise to moiré patterns. In both of these instances the misfitting interfaces were normal to the electron beam and parallel to substantially parallel sided faces of the crystals. In other cases fringe patterns have been observed in slightly misoriented boundaries in metal films inclined at an angle to the electron beam. Here there may be complications due to the pattern being formed in two overlapping wedge shaped crystals and for a full interpretation effects due to extinction (§ 4.2.4) must be taken into account.

§ 6. Conclusions

It has been shown that there are two basic approaches to the electron microscopic observation of detail of molecular and atomic dimensions in crystals. Surface methods are now sufficiently refined for special features such as steps to be observable right down to atomic dimensions, and there is undoubtedly much valuable information still to be obtained from the study of such surface features. However, where in the past surface methods had been used in attempts to infer internal structure from observations of surface features because of inadequacy of resolution or insufficient refinement of specimen preparation methods, it is likely that in the future there will be an increasing interest in direct transmission methods. With these it is possible to utilize fully the enhanced resolution of the lens for periodic objects and also to supplement the information obtained from the micrograph with that obtainable from the diffraction pattern. In transmission the fundamental periodicity of the crystal lattice may be directly resolved by the moiré technique. In both the direct and indirect method, the kinematic theory of diffraction is able to give a reasonably simple interpretation of many of the effects observed. It is clear, however, that there is a need for a more refined dynamical treatment of the problems of image formation.

From the practical point of view, in applying moiré methods it is important to know what is the maximum permissible physical separation of the two specimens for the relevant beams still to be coherent so that interference phenomena are observable. There is little experimental data available on this at the moment. Rang (1953) has observed what are essentially moiré effects between two lamellae of mica forming a 'bubble' in mica, where the separation of the two mica flakes was ~ 460 Å. Mollenstedt and Duker (1953) observed long range interference effects in overlapping crystals of molybdenum sulphide where the path difference

between the interfering beams was 84 wavelengths. These experiments indicate that close physical contact of the two specimens is not essential. This is of some importance since one can envisage experiments in which it is required to use one crystal as a fixed reference grating while some change is observed in the other. The use of epitaxial growth phenomena for the production of one oriented phase upon another produces moiré patterns which certainly have a great intrinsic interest from the point of view of oriented growth, but to avoid the special problems arising from the growth mechanism itself it is advisable to use the technique of superposition of separately prepared crystals for the proper study of the properties of one of them.

Acknowledgments

I wish to thank those acknowledged in the captions who have made available prints of micrographs for reproduction, my colleagues Mr. G. A. Bassett and Dr. D. W. Pashley for permitting me to refer to unpublished work, and Dr. D. W. Pashley for his helpful comments on the manuscript.

The paper is published by permission from the Chairman of Tube Investments Ltd.

References

Abbe, E., 1837, *Arch. mikr. Anat.*, **9**, 413.

Anderson, N. G., and Dawson, I. M., 1953, *Proc. roy. Soc.* A, **218**, 255 ; 1955, *Ibid.*, **228**, 539.

Barrer, R. M., 1956, *Physical chemistry of some non-stoichimetric phases* (Brussels : Solvay Congress).

Barrer, R. M., Bultitude, F. W., and Sutherland, J. W., 1957, *Trans. Faraday Soc.*, **53**, 1111.

Bassett, G. A., 1958 (in the press).

Bassett, G. A., and Menter, J. W., 1957, *Phil. Mag.*, **2**, 1482.

Bassett, G. A., Menter, J. W., and Pashley, D. W., 1958, *Proc. roy. Soc.* (in the press).

Bergerhof, G., Koyama, H., and Nowacki, W., 1956, *Experentia*, **12**, 418.

Bernard, R., and Pernoux, E., 1953, *C.R. Acad. Sci., Paris*, **236**, 187.

Boersch, H., 1947, *Z. Naturf.*, **2 a**, 615.

Bradley, D. E., 1954, *Brit. J. appl. Phys.*, **5**, 65 ; 1958, *Nature, Lond.*, **181**, 875.

Brindley, G. W., Comer, J. J., Uyeda, R., and Zussman, J., 1958, *Acta. cryst.*, **11**, 99.

Burton, W. K., Cabrera, N., and Frank, F. C., 1951, *Phil. Trans. roy. Soc.* A, **243**, 299.

Cochran, W., and Kartha, G., 1956 a, *Acta cryst.*, **9**, 259 ; 1956 b, *Ibid.*, **9**, 941 ; 1956 c, *Ibid.*, **9**, 944.

Conrady, A. E., 1919, *Mon. Not. R. astr. Soc.*, p. 575 (June).

Cowley, J. M., and Moodie, A. F., 1957 a, *Proc. phys. Soc. Lond.* B, **72**, 486 ; 1957 b, *Ibid.*, **72**, 497 ; 1957 c, *Ibid.*, **72**, 505.

Cowley, J. M., Rees., A. L. G., and Spink, J. A., 1951, *Proc. phys. Soc. Lond.* A, **64**, 609.

Darbyshire, J. A., and Cooper, E. R., 1935, *Proc. roy. Soc.* A, **15**, 104.

Dawson, I. M., 1951, *Nature, Lond.*, **168**, 241 ; 1952, *Proc. roy. Soc.* A, **214**, 72.

Dawson, I. M., and Vand, V., 1951 a, *Nature, Lond.*, **167**, 476 ; 1951 b, *Proc. roy. Soc.* A, **206**, 555.

Dowell, W. C. I., Farrant, J. L., and Rees, A. L. G., 1956, *Proc. of the 3rd International Conference on Electron Microscopy, London*, 1954, p. 279 ; 1957, *Electron microscopy. Proc. First Regional Conf. in Asia and Oceania, Tokyo*, 1956, p. 320.
Drechsler, M., 1956, *Z. phys. Chem.*, **6**, 272.
Drechsler, M., Pankow, G., and Vanselow, R., 1955, *Z. phys. Chem.*, **4**, 249.
Ennos, A. E., 1953, *Brit. J. appl. Phys.*, **4**, 101 ; 1954, *Ibid.*, **5**, 27.
Farrant, J. L., and Hodge, A. J., 1956, *Proc. of the 3rd International Conference on Electron Microscopy, London*, 1954, p. 118.
Farrant, J. L., and Rees, A. L. G., 1956, Paper to International Union of Crystallography Symposium on *Structures on a scale between atomic and molecular dimensions*, Madrid, April 2–7.
Fernandez-Moran, H., 1956, *Industr. Diam. Rev.*, **16**, 128.
Finch, G. I., and Sun, C. H., 1936, *Trans. Faraday Soc.*, **32**, 852.
Finch, G. I., and Wilman, H., 1937, *Ergebn. exakt. Naturw.*, **16**, 353.
Forty, A. J., 1954, *Advanc. Phys.*, **3**, 1.
Frank, F. C., 1951, *Phil. Mag.*, **42**, 1014.
Gabor, D., 1949, *Proc. roy. Soc.* A, **197**, 454.
Gard, J. A., 1956, *Ph.D. Thesis*, University of Aberdeen.
Gard, J. A., Barrer, R. M., and Baynham, J., 1955, *J. chem. Soc.*, 2480.
Gard, J. A., and Taylor, H. F. W., 1957, *Miner. Mag.*, **31**, 361.
Good, R. H., and Müller, E. W., 1956, *Handbuch der Physik*, **21**, 174.
Goodman, J. F., 1957, *Nature, Lond.*, **180**, 425.
Green, T. A., and Weigle, J., 1948, *Helv. phys. acta*, **21**, 217.
Guild, J., 1956, *The interference systems of crossed diffraction gratings* (Oxford : Clarendon Press).
Haanstra, H. B., 1955, *Philips tech. Rev.*, **17**, 178.
Haine, M. E., 1954, *Advanc. Electron.* **6**, 295 ; 1957, *J. sci. Instrum.*, **34**, 9.
Haine, M. E., and Jervis, M. W., 1954, *J. Instn. elect. Engrs.*, **102 B**, 265.
Haine, M. E., and Mulvey, T., 1952, *J. Opt. Soc. Amer.*, **42**, 763 ; 1954, *J. sci. Instrum.*, **31**, 326.
Hall, C. E., 1951, *J. biol. Chem.*, **185**, 45.
Hamm, F. A., and van Norman, E., 1948, *J. appl. Phys.*, **19**, 1097.
Hashimoto, H., and Uyeda, R., *Acta. cryst.*, **10**, 143.
Heidenreich, R. D., 1942, *Phys. Rev.*, **62**, 291.
Hillier, J., 1941, *Phys. Rev.*, **60**, 742 ; 1954, *Nat. Bur. Stand. Circ.*, No. 527, " Electron Physics ", 413.
Hillier, J., and Ramberg, E., 1947, *J. appl. Phys.*, **18**, 48.
Hirsch, P. B., 1956, *Proc. Metal Phys*, **6**, 236.
Hirsch, P. B., Horne, R. W., and Whelan, M. C., 1956, *Phil. Mag.*, **1**, 677.
Hopkins, H. H., 1953, *Proc. roy. Soc.* A, **217**, 408.
Karasaki, S., and Komoda, T., 1958, *Nature, Lond.*, **181**, 407.
Keith, H. D., 1956, *Proc. Phys. Soc. Lond.* B, **69**, 180.
Keller, A., and O'Connor, A., 1958, Faraday Soc. Discussion on *Configurations and interactions of macromolecules and Liquid crystals*, Leeds.
Kinder, E., 1943, *Naturwissenschaften*, **31**, 149.
Koch, and König, 1954, Münster meeting of German E.M. Society.
Komoda, T., Suito, E., Uyeda, N., and Watanabe, H., 1958, *Nature, Lond*, **181**, 332.
Kossel, W., 1943, *Naturwissenchaften*, **31**, 323.
Labaw, L. W., and Wyckoff, R. W. G.,1955, *Exp. Cell Res. Suppl.*, **3**, 395 ; 1956 a, *Science*, **123**, 849 ; 1956 b, *Proc. Kon. Ned. Akad. Met.*, **59 B**, 451 ; 1957 a, *Arch. Biochem. Biophys.*, **67**, 225 ; 1957 b, *Proc. nat. Acad. Sci. Wash.*, **43**, 1032.

Leisegang, S., 1956 a, *Handbuch der Physik*, **33,** 396 ; 1956 b, *Proceedings 3rd International Conference on Electron Microscopy, London*, 1954, 184.
Le Poole, J. B., 1956, Colloques Internationaux du C.N.R.S., Toulouse, 1955. *Les techniques récentes en microscopie électronique et corpusculaire*, 142.
Locquin, M., 1956, Colloques Internationaux du C.N.R.S., Toulouse, 1955. *Les techniques récentes en microscopie électronique et corpusculaire*, 142.
Marton, L., and Schiff, L. I., 1941, *J. appl. Phys.*, **12,** 759.
MacMullan, D., 1952, *Proc. Inst. elect. Engrs*, **100,** 245.
Menter, J. W., 1952, *J. Inst. Metals*, **81,** 163 ; 1953, *J. Photo. Sci.*, **1,** 12 ; 1956 a, *Proc. roy. Soc.* A, **236,** 119 ; 1956 b, *Times Sci. Rev.*, No. 22, 17 ; 1957, Electron microscopy. Proceedings of Stockholm Conference, 1956, 88.
Mitsuishi, T., Nagasaki, H., and Uyeda, R., 1951, *Proc. imp. Acad. Japan*, **27,** 86.
Möllenstedt, G., 1956, *Proceedings 3rd International Conference on Electron Microscopy, London*, 1954, 694.
Möllenstedt, G., and Düker, H., 1953, *Phys. Verh.*, **4,** 98.
Müller, E. W., 1951, *Z. Phys.*, **131,** 136 ; 1956, *J. appl. Phys.*, **27,** 474 ; 1957, *Ibid.*, **28,** 1.
Neider, R., 1957, *Electron Microscopy. Proceedings of Stockholm Conference*, 1956, 93.
Niehrs, H., 1956, *Optik*, **13,** 399 ; 1957, *Electron microscopy. Proceedings of Stockholm Conference*, 1956, 86.
Pashley, D. W., 1951, *Proc. roy Soc.* A, **210,** 354 ; 1958 (in the press).
Pashley, D. W., Menter, J. W., and Bassett, G. A., 1957, *Nature, Lond.*, **179,** 752.
Paterson, M. S., 1952, *J. appl. Phys.*, **23,** 805.
Pease, R. S., 1952, *Acta. cryst.*, **5,** 356.
Porter, A. B., 1906, *Phil. Mag.*, **11,** 154.
Price, W. C., and Wyckoff, R. W. G., 1946, *Nature, Lond.*, **157,** 764.
Raether, H., 1932, *Z. Phys.*, **78,** 527.
Rang, O., 1953, *Z. Phys.*, **136,** 465.
Rayleigh, Lord, 1874 a, *Phil. Mag.*, **47,** 81 ; 1874 b, *Ibid.*, **47,** 193.
Riley, D. P., 1951, *Nature, Lond.*, **168,** 241.
Robertson, J. M., 1934, *J. chem. Soc.*, 615.
Robertson, J. M., and Woodward, I., 1940, *J. chem. Soc.* 36.
Scherzer, O., 1947, *Optik*, **2,** 114 ; 1949 a, *Ibid.*, **5,** 497 ; 1949 b, *J. appl. Phys.*, **20,** 20.
Schiff, I. J., 1942, *Phys. Rev.*, **61,** 721.
Schott, O., and Leisegang, S., 1957, *Electron Microscopy. Proceedings of Stockholm Conference* 1956, p. 20.
Seeliger, R., 1953, *Optik*, **10,** 29.
Seki, Y., 1951, *J. Phys. Soc. Japan*, **6,** 534, 1951 ; 1953, *Ibid.*, **8,** 149.
Smith, K. C. A., and Oatley, C. W., 1955, *Brit. J. appl. Phys.*, **6,** 391.
Suito, E., and Uyeda, N., 1957, *Proc. imp. Acad. Japan*, **33,** 398.
Suzuki, T., 1957, Private communication from B. T. M. Willis.
Taylor-Jones, E., 1934, *Phil. Mag.*, **18,** 291.
Whelan, M. W., and Hirsch, P. B., 1957 a, *Phil. Mag.*, **2,** 1121 ; 1957 b, *Ibid.*, **2,** 1303.
Whelan, M. W., Hirsch, P. B., Horne, R. W., and Bollman, W., 1957, *Proc. Roy. Soc.* A, **240,** 524.
Willis, B. T. M., 1957, *Proc. roy. Soc.* A, **239,** 192.
Wilman, H., 1940, *Proc. phys. Soc. Lond.*, **52,** 325.
Wilson, A. J. C., 1949, *Research. Lond.*, **2,** 541.
Wyckoff, R. W. G., 1956, Colloques Internationaux du C.N.R.S., Toulouse, 1955, *Les techniques récentes en microscopie électronique et corpusculaire*, 135.

Dielectric Breakdown in Solids

By J. J. O'Dwyer†

Commonwealth Scientific and Industrial Research Organization
National Standards Laboratory,
Division of Electrotechnology,
City Road, Chippendale, N.S.W., Australia

Contents

---

† Present address: The New South Wales University of Technology, School of Applied Physics, Kensington, N.S.W., P.O. Box 1.

§ 1. Introduction

1.1. *The Breakdown of Dielectric Solids in Large Fields*

The types of breakdown most commonly met in engineering practice are different in nature from those occurring in pure materials under controlled laboratory conditions.† These latter types of breakdown are naturally of greater interest in physics.

The most frequent breakdowns encountered in electrical engineering are those due to discharges in the external medium, or in some gap or imperfection within the insulator itself. In many cases the insulator can be destroyed as a result of either of these types of discharge. For the special case in which the initial discharge is along the surface of the insulator, this type of breakdown is referred to as high voltage tracking.

Another manner of destruction of the insulating properties of a dielectric is called thermal breakdown. It amounts to melting of the dielectric due to its inability to dissipate the heat generated by the current flow in the electric field. The main distinguishing characteristics of thermal breakdown are :

(i) It requires a time of at least milleseconds to occur, and in many cases very much longer. This means that thermal breakdown cannot be observed at all with impulse voltages of very short duration.

(ii) The dielectric strength depends on the size and shape of the sample and the electrode configuration. Arrangements in which heat can be more rapidly conducted away will yield higher breakdown strengths for the same material.

(iii) For alternating fields of high frequency the breakdown strength will usually be lower than for the same arrangement of insulator and electrodes in a d.c. field. This occurs since the total power loss in a dielectric usually increases with increasing frequency.

Similar to what has been usually known as thermal breakdown (in fact it may be considered as a limiting case of thermal breakdown) is a manner of breakdown which has been observed in some materials at high temperatures, where both d.c. and impulse tests have been done. Its distinguishing features are :

(i) The breakdown strength varies markedly with the duration of the applied voltage—this is the chief characteristic of this type of breakdown.

(ii) The dielectric strength does not depend on the size of the sample and the electrode configuration, provided that the arrangement is such that heat is not conducted away too rapidly.

These two classes of thermal breakdown will be more adequately distinguished below.

---

† For a full account of the types of dielectric breakdown in solids commonly met in practice see, for example, Whitehead (1950).

A third type of dielectric breakdown which has been observed has been referred to as intrinsic breakdown. In a general way it can be said that the main features of intrinsic breakdown are :

(i) It occurs in a time of order of microseconds or less.

(ii) Within wide limits the breakdown strength does not depend on the size and shape of the sample, or on the material or configuration of the electrodes. On this account it is regarded as a characteristic constant of the dielectric at a given temperature, and the process of breakdown is referred to as intrinsic breakdown.

Closely related to intrinsic breakdown, but still distinct from it, is what has been called avalanche breakdown. The features which distinguish it experimentally are similar to those for intrinsic breakdown, but with the very important exception that the breakdown strength must depend on the thickness of the dielectric if avalanche breakdown occurs.

The main purpose of this article is to give as coherent as possible an account of the theories of intrinsic breakdown of solids, with particular emphasis on the physical process envisaged by each theory. Theories of the modified form of thermal breakdown which occurs in impulse testing will also be discussed.

1.2. *Measurements of Breakdown Strength*

It is of interest to set out the general experimental facts of breakdown, found under controlled laboratory conditions, which the theories are called upon to explain. A somewhat more detailed reference to experimental work is left over until it can be compared with the results of various theories.

For reasons which will become apparent later it is convenient to divide dielectric solids into three main groups when breakdown properties are being discussed. These are :

(i) Polar crystals.

(ii) Non-polar crystals.

(iii) Amorphous substances.

The principal insulating polar crystals are the alkali halides, mica and crystalline quartz which have all been the subject of extensive investigation. Non-polar crystals would include for example diamond and long chain paraffins in the crystalline solid state, but there does not appear to be much experimental work available on these substances. Under the heading of amorphous substances we have glass, fused quartz and most of the polymers. This last group has been well covered by experiment.

Considering first the alkali halides, we find from a study of experimental papers on the subject certain agreements and disagreements between various authors (cf. Buehl and von Hippel 1939, Austen and Whitehead 1940, von Hippel and Alger 1949, Calderwood and Cooper 1953, and Konorova and Sorokina 1957). It is fairly generally agreed (at least for

the chlorides and bromides of sodium and potassium) that a transition temperature exists at about 50°c above and below which the breakdown mechanisms differ. It has been customary then to distinguish for the alkali halides both a low and a high temperature region of breakdown, corresponding to temperatures respectively below and above this transition temperature.

In the low temperature region there is considerable agreement among experimental workers. It is found that :

(i) The breakdown strength is of the order of 10^6 volts per cm and increases slowly with temperature from liquid air temperature up to the transition point.

(ii) Impulse testing with variation of the pulse time from seconds down to 10^{-8} sec shows that the breakdown strength is independent of the pulse time within this range.

(iii) The breakdown strength is independent of thickness roughly over the region 10^{-3} to 10^{-1} cm. These last two notes are characteristic of what we have called intrinsic breakdown, and it must be considered as established that this type of breakdown occurs in the alkali halides over the range of experimental conditions described.

On the other hand there is conflicting evidence in the high temperature region for these substances. Broadly speaking, there are two positions :

(i) That the breakdown strength falls slowly with increasing temperature above the transition temperature and its value is independent of whether it is measured by d.c. or by pulses longer than a microsecond duration.

(ii) That the fall in breakdown strength occurs with d.c. tests, but with microsecond pulses the breakdown strength rises slowly with increasing temperature in such a way as to represent a continuation of its behaviour below the transition temperature. For longer pulses the values obtained are between those for these two extremes. These alternative opinions will be discussed in more detail later, but it is clear that the occurrence of intrinsic breakdown is not established in the high temperature region of the alkali halides.

Mica has been the subject of extensive experimental work (cf. Austen and Whitehead (1940) and Kawamura and co-workers (1953, 1954). The different types of mica show small measured differences but the general features are quite clear. They are :

(i) The breakdown strength is of the order of 10^7 volts per cm and varies very little with temperature over a range of 200°c either side of room temperature.

(ii) There is no difference between the results of d.c. and impulse tests over the usual range of rise times.

(iii) The breakdown strength is independent of the thickness over the range 10^{-4} to 5×10^{-2} cm.

(iv) As the thickness decreases below 10^{-4} cm the breakdown strength rises rapidly. It will be seen that the breakdown of mica fulfills the concept of intrinsic over a wide range, but for very thin flakes there appears to be at least a modification in the mechanism of breakdown.

Amongst amorphous substances glass has received special attention (cf. Keller 1951 and Vermeer 1954,1956). These authors have used test voltages rising linearly with time, and have done measurements over a complete range of rise times from minutes to microseconds. There are small differences for the various types of glass but the general conclusions are :

(i) The breakdown strength is of the order of 10^7 volts per cm and does not vary with temperature for sufficiently low temperatures and short rise times.

(ii) A transition temperature exists above which the breakdown strength decreases with increasing temperature. This transition temperature is higher for shorter voltage rise times.

(iii) The breakdown strength is independent of thickness over the range 10^{-4} to 10^{-2} cm. Under these experimental conditions the breakdown of glass is intrinsic below the transition temperature, and is a type of thermal breakdown above it.

The experimental evidence produced above, particularly the constancy of breakdown strength over pulse times down to 10^{-8} sec in some cases, indicates that intrinsic breakdown is an electronic process. Experiments to support this idea were performed by von Hippel (1935) and von Hippel and Lee (1941) who found that the addition of small amounts of foreign ions to an ionic crystal always raised the breakdown strength. Granted the electronic nature of intrinsic breakdown this is explained quite naturally in a way analogous to that in which the increase in electrical resistance of metals upon addition of foreign atoms is explained.

A certain amount of work has been done on other solids, some of which will be mentioned later in detailed comparison with experiment. However, the work cited on the alkali halides, mica and glass gives typical instances of the facts of intrinsic breakdown which theories are called upon to explain.

1.3. *The General Criteria for Breakdown*

1.3.1. *Intrinsic breakdown*

When an electric field is applied to an insulator conduction electrons will gain energy from the field at a rate

$$A = jF \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where j is the electric current and F is the applied field. This relation holds whether any steady condition has been reached or not. If in addition a mechanism exists for transferring electron energy to the lattice (e.g. inelastic collisions between electrons and lattice vibrations) and we denote the rate of this energy transfer by B, then the condition for a

steady state is $$A = B \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

in so far as processes affecting the electrons are concerned. In general both A and B will be complicated functions of F, but such that the condition (2) can be satisfied for values of F up to a certain critical value F^* and not beyond it. This critical value is the intrinsic breakdown strength for the particular transfer mechanism being considered according to the ideas of von Hippel (1935) and Fröhlich (1937, 1947, 1956). Various criteria for intrinsic breakdown arise from a consideration of the different ways in which different types of energy transfer from electrons to lattice can establish a steady state inasmuch as the electrons are concerned. Thus intrinsic breakdown clearly relates to the failure of a steady state situation. Since this steady state involves only electronic processes then the breakdown strength will be independent of the time of application of the electric field. This, as already mentioned, is one of the principal distinguishing features of this type of breakdown.

For intrinsic breakdown such notions as the specific heat or the thermal conductivity do not enter explicitly, since it is assumed that the electrical conductivity is sufficiently low and the thermal conductivity sufficiently high that the power generated by the electric current is conducted away with only a very small temperature rise. The breakdown will then be independent of the size and shape of the dielectric specimen and the thermal conductivity of the electrodes, which is the second main distinguishing feature of intrinsic breakdown. In this connection it is interesting to note that the alkali halides are ideal substances in which to seek intrinsic breakdown since in good crystals the electrical conductivity decreases and the thermal conductivity increases with decreasing temperature. We could thus conclude on these grounds that intrinsic breakdown should occur in alkali halides for sufficiently low temperatures. We have seen that this is, in fact, the case. These considerations do not preclude amorphous dielectrics from exhibiting a region of intrinsic breakdown. Although the thermal conductivity decreases with decreasing temperature in most such cases, nevertheless, it decreases more slowly than does the electrical conductivity so that again it is reasonable to look for a region of intrinsic breakdown at low temperatures.

It should not be thought that intrinsic breakdown conceives of a process in which the final destruction of the lattice is explained by electronic instability alone. Obviously if the material is destroyed, then at some stage the properties of the lattice will enter. However it is asserted that the initial irreversible instability is electronic in nature, and that in cases in which intrinsic breakdown is observed this irreversible electronic catastrophe produces conditions in which the lattice is destroyed.

1.3.2. *Thermal breakdown*

Consider a dielectric under the action of an applied field in which we can write the energy transfer to the lattice as

$$B = \sigma F^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where σ is the electrical conductivity. This implies that we always have at least a quasi steady state as far as the electronic processes are concerned. By this we mean that if the applied field varies with time then this variation must be sufficiently slow that electronic processes have assumed the steady state condition appropriate to the field strength at that time. The energy transfer to the lattice will partly go to increase the lattice temperature and partly be conducted away. Hence we have

$$C_v \frac{dT}{dt} + \operatorname{div}(\kappa \operatorname{grad} T) = \sigma F^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where C_v is the specific heat per unit volume, κ the thermal conductivity, and dT/dt is the time derivative and grad T the space gradient of the temperature. Equation (4) is the most general form for the fundamental equation of thermal breakdown. Since C_v, κ and σ may all be functions of T and σ is also a function of F the equation cannot be solved in general terms, but the object of a solution would be to calculate the temperature at all points of the dielectric as a function of time for a given applied field strength and with suitable boundary conditions. Assuming that the dielectric is destroyed if some temperature T_m (not necessarily the melting point of the dielectric) is attained at any point of it, then this solution gives the time required by a given applied field to destroy the material. Numerical solutions have been carried out for some materials (cf. Whitehead 1950) and the general features of these solutions are shown in fig. 1. They are :

(i) For low applied field strengths the temperature T_m is not attained in the dielectric and there is no breakdown of insulating properties.

(ii) For a critical field strength F_m the temperature T_m is attained asymptotically after an infinite time. This field strength is the minimum thermal breakdown voltage.

(iii) For field strengths greater than F_m the temperature is attained in a finite time which depends on the field. These field strengths are the thermal breakdown field strengths as a function of time of application of the field. In case (*a*) the thermal conduction term of eqn. (4) remains important, while in case (*b*) it is negligible on account of the short time taken to exceed the critical temperature.

Bearing in mind these features of the general solution two limiting cases arise which have been treated extensively. The first case arises from the recognition that conditions for the minimum thermal breakdown voltage amount to a steady state for the lattice processes with the temperature of the hottest part of the dielectric put equal to T_m. On account of the steady state the time differential is put equal to zero and eqn. (4) is then solved for F_m. This is the usual calculation of thermal breakdown strength and it has been fully reviewed by Whitehead (1950).

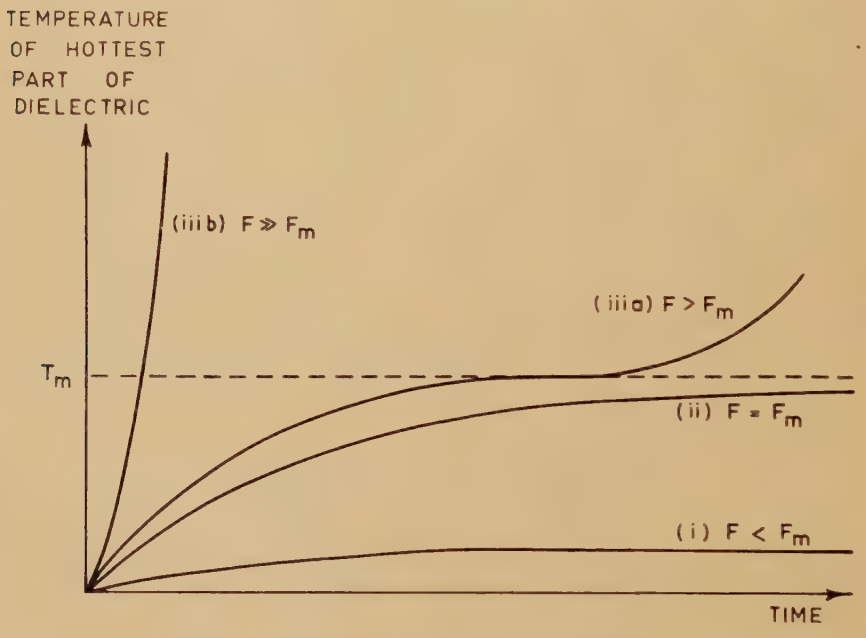
The other limiting case arises when the time of application of the electric field is very small—of the order of seconds and less. Under these

circumstances it is often permissible to ignore the heat conduction term in eqn. (4). The simplified equation

$$C_v \frac{dT}{dt} = \sigma F^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

can then be used to determine the critical field strength as a function of time. We shall call these field strengths the impulse thermal breakdown strengths to bring out the point that they are functions of the time of application of the voltage impulse.

Fig. 1



Schematic diagram of solutions to equation (4).

This approach has been used by Franz (1952, 1956) in his theories of breakdown. Franz has calculated the variation of σ with T and F for various mechanisms which cause an increase in the electronic conductivity at high field strengths. This calculated σ can then be used in (5) to determine impulse thermal breakdown strengths. The same method was used in a more phenomenological way by Vermeer (1956) to give a theory of the breakdown of glass in a certain region. In this work the dependence of σ on T and F was found by experiment.

Impulse thermal breakdown will be discussed below since there is some evidence that it occurs in various substances when the temperature is too high to favour intrinsic breakdown.

1.3.3. *Avalanche breakdown*

A criterion for breakdown has been put forward by Seitz (1949) who has examined the formation of avalanches of conduction electrons, and

has assumed that breakdown occurs when these avalanches exceed a certain critical size.

The critical size for an avalanche is obtained by an order of magnitude calculation. Thus we assume that the avalanche is to occur in a cylindrical region 1 cm long and 10^{-3} cm in radius. For a field of order of 10^6 volts per cm each electron would receive 10^6 ev of energy in this path length. If an energy of 10 ev per atom is considered sufficient to destroy the insulator, then there must be one electron in the avalanche for each 10^5 atoms in the cylindrical region. But such a cylindrical region will contain about 10^{17} atoms which immediately gives the value 10^{12} as the critical number of electrons in an avalanche.

Suppose now that the avalanche is built up by one free electron which produces a secondary ; this pair produce two more and so on. Then the number of generations required to produce the critical sized avalanche is given by $2^n = 10^{12}$ or $n = 40$. This means that in order to produce the critical-sized avalanche the interelectrode distance must be forty times the mean free path for ionization at that field strength.

These considerations show that if avalanche breakdown is to be observed it will possess the following characteristics :

(i) The pre-breakdown current will be very noisy due to the build-up of avalanches which fail to reach the critical size.

(ii) The breakdown strength will be a function of interelectrode spacing.

Some evidence that breakdown of this nature occurs in mica under certain conditions is given below.

§ 2. The Interaction between a Conduction Electron and Lattice Vibrations

2.1. *The Electronic Structure of Dielectrics*

In solids in general the electrons occupy zones or bands in energy space. The more closely bound the electron the smaller is the region of energy of the allowed band, and the larger is the region of energy of the forbidden band separating it from the next allowed band. Thus the simple picture of an insulator at absolute zero offered by quantum mechanics is one in which bands are completely occupied by electrons up to a certain energy, and completely unoccupied thereafter. A fully occupied energy band cannot conduct since conduction requires empty levels into which electrons can be accelerated. Neither of course can an unoccupied allowed band conduct, so that such a model is perfectly insulating.

At temperatures other than the absolute zero all real insulators differ from this ideal model in two important aspects. In the first place the lowest unoccupied band will always contain some electrons ionized from lower levels. This band is referred to as the conduction band of the insulator. Because of the great number of vacant levels all the electrons in it can be accelerated by an applied field, and the exclusion principle is not of importance. The second modification to the ideal model arises

from the presence of electron energy levels other than those in the bands. The most important of these for breakdown considerations are localized energy levels lying just below the conduction band. Since they are localized, electrons in them do not conduct, but their importance lies in the fact that ionization from them to conduction levels and recombination to them from conduction levels may be much more probable than the same type of processes between conduction levels and the energy levels in the next lower band. Generally speaking we can, at low temperatures, consider a dielectric as being well represented by the ideal model with the addition of some electrons in conduction levels. It is usually only at higher temperatures (room temperature or higher) that the consideration of isolated energy levels appears to be of any importance to breakdown theory.

2.2. *Inelastic Collisions with Lattice Vibrations*

In all solids the acceleration of an electron is not hindered by a perfectly periodic field ; the resistance to their motion arises from interaction between the electron and the lattice vibrations leading to an energy transfer. Any theory of electronic breakdown in solids, whether it be a theory of intrinsic breakdown or impulse thermal breakdown depends on a theory of this energy transfer. Such a theory being a big subject in itself, with applications in many other fields besides breakdown, will accordingly be treated briefly and moreover oriented explicitly to the requirements of breakdown theory.

Fröhlich and Seitz (1950) have discussed the energy of interaction of an electron with lattice vibrations in non-metals, and have shown that it can be split into three principal terms :

(i) Interaction with a dipole field which is derived by treating the ions of an ionic crystal as point charges. A longitudinal mode of vibration in such a model is a longitudinal polarization wave. Any distortion of electron shells is ignored.

(ii) Interaction with the dipole component of the field arising from the distortion of the electron shells. This is closely related to (i).

(iii) Interaction with the non-dipolar component of the field arising from the distortion of the electron shells. All three interactions occur in polar crystals, but the last component is the only one that occurs in non-polar insulators.

Taking the case of ionic crystals and considering only the optical mode of vibration, the total interaction energy has been determined in two principal manners each with its own region of validity.† One method (cf. Fröhlich 1937) gives the interaction in terms of the microscopic physical constants of the lattice, namely the ionic charge, the ionic masses

† Some more complicated treatments of the interaction have been given (cf. Chuenkov 1954 and Veelken 1955) in which the effect of accoustical modes is also treated.

and the interionic distance. In this method the interaction (i) is treated as the only one on the assumption that the combined effect of (ii) and (iii) is negligible. This has in fact been shown to be a reasonable approximation when the most important processes are due to the short polarization waves of the lattice.

Another method (cf. Fröhlich, Pelzer and Zienau 1950) treats the polarization waves as occurring in a continuum, and hence states the interaction in terms of the macroscopic physical constants of the lattice—the high and low frequency dielectric constants and the lattice vibration frequency. In this way interactions (i) and (ii) are taken into account. It is of interest to note that another calculation of the interaction (cf. Callen 1949), in which the lattice is treated as a set of point charges having a certain effective charge differing from the ionic charge, is equivalent to the continuum model although the procedure of the calculation is different. The reason for this is that the effective ionic charge is determined from macroscopic considerations. It is evident that the continuum approximation is a very good one in any case in which the processes of importance are due to an interaction with the long polarization waves of the lattice.

Since in most of the applications to polar crystals which follow the continuum approximation is the appropriate one, we shall use it for the purpose of uniformity in the presentation of results.

The interaction energy for non-polar crystals is determined by methods similar to those used in the conduction theory of metals.

2.3. *The Perturbation Method Applied to Ionic Crystals*

Consider an electron interacting with a continuous dielectric medium having a single rotational frequency ω for longitudinal polarization waves, and low and high frequency dielectric constants ϵ_0 and ϵ_∞.

Following the method given by Fröhlich (1954) we divide the Hamiltonian into three parts

$$H = H_{\text{osc.}} + H_{\text{El.}} + H_{\text{Int.}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

in which the medium is represented by a set of harmonic oscillators, the electron by a particle of mass m and charge e, and the interaction energy is put as the energy of this particle in the field produced by the polarization waves. The electron interacts only with longitudinal waves which produce a polarization

$$\mathbf{P}(\mathbf{r}) = (1/\sqrt{V}) \sum_{\mathbf{w}} \{\mathbf{a}_{\mathbf{w}} \exp(i\mathbf{w}\cdot\mathbf{r}) + \mathbf{a}_{\mathbf{w}}^* \exp(-i\mathbf{w}\cdot\mathbf{r})\}. \quad . \quad . \quad (7)$$

For free oscillations the $\mathbf{a}_{\mathbf{w}}$ are periodic in time, and we have taken the angular frequency ω to be independent of w. Introducing the electric potential $\phi(\mathbf{r})$ related to $\mathbf{P}(\mathbf{r})$ by

$$4\pi\mathbf{P}(\mathbf{r}) = \operatorname{grad}\phi(\mathbf{r}) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and new coordinates

$$\left.\begin{aligned} X_{\mathbf{w}} &= a_{\mathbf{w}} + a_{\mathbf{w}}^* \\ Y_{\mathbf{w}} &= -iM\omega(a_{\mathbf{w}} - a_{\mathbf{w}}^*) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

we obtain from (8) and (9)

$$\phi(\mathbf{r}) = (4\pi/\sqrt{V})\sum_{\mathbf{w}} \frac{1}{w}\{X_{\mathbf{w}} \sin \mathbf{w}\cdot\mathbf{r} + (Y_{\mathbf{w}}/M\omega)\cos \mathbf{w}\cdot\mathbf{r}\}. \quad . \quad (10)$$

The medium can be represented by harmonic oscillators and we choose the constant M so that

$$\left.\begin{aligned} H_{\text{osc.}} &= 2M\omega^2 \sum_{\mathbf{w}} a_{\mathbf{w}} a_{\mathbf{w}}^* \\ &= \tfrac{1}{2}\sum_{\mathbf{w}} \{M\omega^2 X_{\mathbf{w}}^2 + Y_{\mathbf{w}}^2/M\} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (11)$$

Also

$$\left.\begin{aligned} H_{\text{El.}} &= \mathbf{p}^2/2m \\ H_{\text{Int.}} &= e\phi(\mathbf{r}_e) \end{aligned}\right\}$$

where $\mathbf{r}_e$ and $\mathbf{p}$ are the position and momentum of the electron. Calculation of the potential due to a point charge in order to determine M then gives

$$\frac{1}{M\omega^2} = \frac{1}{4\pi}\left\{\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

thus determining the interaction energy to be used in the perturbation calculation.

2.3.1. *The rate of energy gain from an applied field*

It is easily shown that the rate of energy gain from an applied field F by a conduction electron whose energy is E may be written

$$A(E, F) = \frac{e^2F^2}{m}\tau(E) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where $\tau(E)$ is the average relaxation time for an electron of energy E.

To calculate the relaxation time we follow Fröhlich (1937) and treat the lattice and the electron quantum mechanically and the interaction as a small perturbation. Thus the unperturbed wave function is

$$\Psi_0 = (1/\sqrt{V})\exp(i\mathbf{k}\cdot\mathbf{r})\prod_{\mathbf{w}} \chi_{\mathbf{w}}(X_{\mathbf{w}}, n_{\mathbf{w}}) \quad . \quad . \quad . \quad . \quad (14)$$

where $\mathbf{k}$ is the electron wave vector and $n_{\mathbf{w}}$ is the number of quanta of the oscillator $\mathbf{w}$. The probability per second of the transition $\mathbf{k}, n_{\mathbf{w}} \rightarrow \mathbf{k}', n'_{\mathbf{w}}$ is then given by

$$P_{\mathbf{w}} = \frac{2\pi}{\hbar}|M|^2\delta(\xi) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

where $\xi = E_{\mathbf{k}'} - E_{\mathbf{k}} + (n_{\mathbf{w}'} - n_{\mathbf{w}})\hbar\omega$ and M is the matrix element of $e\phi$ between the states $\mathbf{k}, n_{\mathbf{w}}$ and $\mathbf{k}, n'_{\mathbf{w}}$. It is found that M vanishes unless momentum is conserved and in the non-vanishing case

$$P_{\mathbf{w}}{}^e = \frac{4\pi^2e^2\omega}{w^2V}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right)(1 + n_{\mathbf{w}})\delta(\xi) \quad . \quad . \quad . \quad . \quad (16)$$

$$P_{\mathbf{w}}{}^a = \frac{4\pi^2e^2\omega}{w^2V}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right)n_{\mathbf{w}}\delta(\xi) \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

for emission or absorption of a quantum $\hbar\omega$ by the electron.

The relaxation time τ is defined by

$$k_x/\tau = -(\partial k_x/\partial t)_{\text{Coll.}}$$
$$= -\sum_{\mathbf{w}} \Delta k_x(\mathbf{w})\{P_{\mathbf{w}}{}^a + P_{\mathbf{w}}{}^e\}. \quad . \quad . \quad . \quad . \quad . \quad (18)$$

where $\Delta k_x(\mathbf{w})$ is the mean change of k_x in one collision with the oscillator $\mathbf{w}$.

Since the transition probabilities do not depend on the azimuth coordinate we can take the average in (18) over this coordinate giving (cf. Fröhlich 1936)

$$\Delta k_x(\mathbf{w}) = -w^2 k_x/2k^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

and replacing the sum in (18) by an integral

$$\frac{1}{\tau} = \frac{V}{(2\pi)^3}\int_0^{2\pi} d\phi \int_{-\pi}^{\pi} \sin\theta\, d\theta$$
$$\times \int_0^{w_0} \frac{w^4}{2k^2}(P_{\mathbf{w}}{}^a + P_{\mathbf{w}}{}^e)\, dw \quad . \quad . \quad . \quad . \quad (20)$$

where w_0 is the upper limit of w for the purpose of this integration, and is determined from different conditions according to the energy of the electron.

The maximum value of the wave number is given by the Debye condition

$$w_{\max} = (6\pi^2)^{1/3}/a \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

and the upper limit of the integral is equal to $w_{\max}$ only if conservation of energy and momentum allow the electron to interact with such short waves. The condition for this is $2k > w_{\max}$ in which case standard methods of integrating (20) give

$$\frac{1}{\tau(E)} = \frac{(6\pi^2)^{2/3}}{8\sqrt{2}} \frac{\hbar e^2 \omega}{a^2 m^{1/2}} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right) \frac{1+2n_{\mathbf{w}}}{E^{3/2}} \quad . \quad . \quad . \quad (22)$$

where $E = \hbar^2 k^2/2m$ is the energy of the electron.

If, on the other hand, we have $2k < w_{\max}$ then the upper limit of the integral arises from energy and momentum considerations alone giving $w_0 = 2k$ and the result is

$$\frac{1}{\tau(E)} = \frac{1+2n_{\mathbf{w}}}{\tau_0(E)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

where

$$\frac{1}{\tau_0(E)} = \left(\frac{m}{2}\right)^{1/2} \frac{e^2\omega}{\hbar} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right) E^{-1/2} \quad . \quad . \quad . \quad . \quad (23\,a)$$

The quantity $\tau_0(E)$ is of importance in later development.

Inasmuch as the limits of integration are concerned the region of validity of (22) is for electron energies in the range

$$\mathscr{I} > E > \hbar^2 w^2_{\max}/8m \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

where $\mathscr{I}$ is the ionization energy, while for (23) we should have

$$\hbar^2 w^2_{\max}/8m > E > 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

In both cases the main contribution to τ comes from the shortest wavelengths included in the integral. In the case of (22) these will be the shortest phonon wavelengths of the crystal and there is no reason to put anything but

$$\omega = \omega_l \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

where ω_l is the Reststrahlen circular frequency. On the other hand, if (23) is used near its lower limit of validity the shortest wavelengths included in the integral will still be among the long phonon waves of the crystal, and we should put

$$\omega = \omega_l(\epsilon_0/\epsilon_\infty)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

(cf. Fröhlich and Mott 1939.) Needless to say, if (23) is used near its upper limit of validity then we again express ω as in (26).

A further restriction on the validity of (22) and (23) arises from the use of perturbation theory to obtain them. As far as (22) is concerned perturbation theory is well justified since the collision rate for such a high energy electron is small compared with ω. On the other hand interaction with short polarization waves of the lattice is predominant so that the continuum approximation is not appropriate, and the interaction energy used in the perturbation calculation is open to objection.† The region of validity of (23) arising solely from the consideration of integration limits would extend down to the lowest energies. However, for energies of order $\hbar\omega$ the collision rate is very high and the use of perturbation theory is unjustifiable. Many theories use the result (23) for electron energies down to several $\hbar\omega$, but the only basis for this extrapolation seems to be that it gives reasonable answers. On the other hand since long waves are primarily responsible for scattering these slow electrons the interaction energy is accurately calculated from the continuum approximation.

Perturbation theory can also be applied to give quantitative results for the scattering of an electron whose energy is less than $\hbar\omega$ interacting with a lattice whose temperature $T \ll \hbar\omega/k_0$. From energetic considerations the electron cannot emit a quantum but only absorb one, and since the lattice temperature is low only a few quanta are present. Accordingly the probability of this absorption is small and using perturbation methods Fröhlich (1954) has found for the average relaxation time of a Maxwell distribution of electrons

$$\frac{1}{\tau} = \left(\frac{2m}{\hbar\omega}\right)^{1/2} \frac{e^2\omega}{\hbar} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right) n_w. \quad . \quad . \quad . \quad . \quad . \quad (28)$$

In this equation ω refers to the longitudinal polarization waves of long wavelength as given by (27).

† For the more appropriate form of the interaction in this case see the original calculation of Fröhlich (1937).

2.3.2. *The energy transfer to the lattice*

The average rate of transfer of energy from the electron to the lattice is given by

$$B(E) = \hbar\omega \sum_{\mathbf{w}} (P_{\mathbf{w}}{}^{e} - P_{\mathbf{w}}{}^{a})$$

$$= \frac{me^2\omega^2}{2\hbar k}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right)\int_{w'}^{w_0} \frac{dw}{w} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

The lower limit w' of the integral† is found from the conservation of energy

$$(\hbar^2k^2 - \hbar^2k'^2)/2m = \hbar\omega$$

which for small w gives

$$w' = m\omega/\hbar k$$

since in this case $k + k' \simeq 2k$ and $k - k' = w'$. As before there will be two possible values for the upper limit of the integral and we find

$$B(E) = \frac{\hbar\omega}{2\tau_0(E)} \log\left\{\frac{(6\pi^2)^{1/2}E^{1/2}}{m^{1/2}a\omega}\right\} \quad . \quad . \quad . \quad . \quad . \quad (30)$$

for the upper range of energy values and

$$B(E) = \frac{\hbar\omega}{2\tau_0(E)} \log\left(\frac{4E}{\hbar\omega}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

for the lower range.

The remarks of the last section regarding the values of ω to be used need to be modified. It is clear that near the bottom of the lower energy range ω should again be determined by eqn. (27). In the remaining energy regions, however, the main contributions to the integral in (29) again come from low values of w, and it is thus reasonable to use the value of ω from (27) in both (30) and (31).

It is of interest to examine the relative effectiveness of long and short polarization waves in the various processes. A summing-up is as follows :

(i) The probability that an electron be inelastically scattered by a phonon is greater for the longer waves (cf. (17)).

(ii) The angle through which an electron is scattered by a phonon is greater for the shorter waves (cf. (19)).

(iii) The main contributions to the relaxation time come from the shorter waves (cf. (20)).

(iv) The main contributions to the energy transfer to the lattice come from the longer waves (cf. (29)).

2.4. *The Perturbation Method Applied to Non-Polar Crystals*

The mobility of an electron in a non-polar insulator has been discussed by Seitz (1948) who has shown that in many cases it is legitimate to treat

† It will be seen that a precise value of the lower limit was not necessary in (16) since $\int_{w'}^{w_0} w\,dw \simeq \int_{w'}^{w_0} w\,dw$ if $w_0 \gg w'$ as is the case.

the interaction between the electron and the lattice as being to the accoustical lattice modes only. Following the usual procedure we adopt a linear relation between the frequency and the wave number of the accoustical modes

$$\omega = ws$$

where s is the sound velocity. The interaction energy is derived in the same manner as in the conduction theory of metals using the deformable atom hypothesis of Bloch. A similar use of perturbation theory to that given for ionic crystals leads to transition probabilities for emission and absorption

$$P_w^{\,e} = \frac{4\pi}{9} C^2 w \frac{\hbar}{M N_a V s} (1 + n_w)\delta(\xi) \quad . \quad . \quad . \quad . \quad . \quad (32)$$

$$P_w^{\,a} = \frac{4\pi}{9} C^2 w \frac{\hbar}{M N_a V s} n_w \delta(\xi). \quad . \quad . \quad . \quad . \quad . \quad (33)$$

where C is the interaction constant of the dimensions of an energy, M is the atomic mass, and N_a the number of atoms per unit volume.

For the case of non-polar crystals we shall be concerned with electrons of thermal energies for which interaction can only be with accoustical modes of low wave number in order to conserve momentum. In this case

$$n_w \simeq k_0 T/\hbar w s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

and the average relaxation time for an electron of energy E is given by

$$\frac{1}{\tau(E)} = \frac{4\sqrt{2}}{9\pi} \frac{C^2 k_0 T M^{3/2}}{\hbar^4 s^2 M N_a} E^{1/2}. \quad . \quad . \quad . \quad . \quad . \quad (35)$$

Considering a Maxwell distribution of electrons at temperature T the average mobility $\mu(T)$ is related to the average relaxation time $\tau(T)$ by

$$\mu(T) = \frac{e}{m} \tau(T). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

Averaging (35) in the appropriate way (cf. Fröhlich 1936), and using the approximate relation

$$s \simeq \{k_0 \Theta\}/\{\hbar (6\pi^2 N_a)^{1/3}\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

where Θ is the Debye temperature we find

$$\left.\begin{aligned} \mu(T) &= \frac{e}{m} \left(\frac{\Theta}{T}\right)^{3/2} \tau_0 \\ \frac{1}{\tau_0} &= (2\pi)^{3/2} \frac{C^2 m^{3/2} s}{\hbar M (k_0 \Theta)^{3/2}} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (38)$$

(cf. Seitz 1948).

The use of perturbation theory is well justified for electrons in non-polar crystals in the regions in which we shall be interested. The form of the interaction energy is, however, somewhat uncertain due to the constant C the evaluation of which is a difficulty in conduction theory. This difficulty need not necessarily arise in breakdown theory if it is possible to state the breakdown strength in terms of the results of conduction experiments, thereby eliminating the interaction constant.

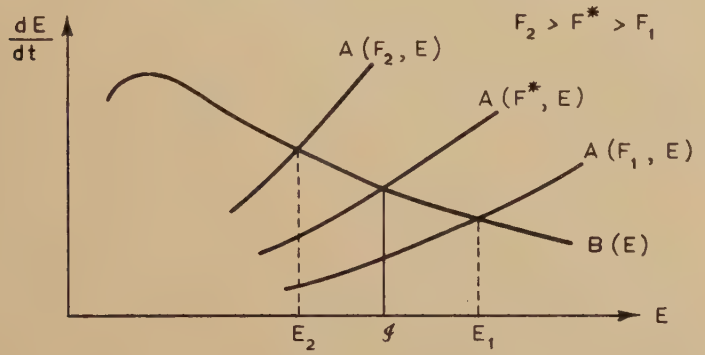
§ 3. Breakdown Theories Involving the Behaviour of one Electron

3.1. *The High Energy Breakdown Criterion for Ionic Crystals*

A theory of breakdown involving the behaviour of one electron does not imply that we are considering the conduction levels of the dielectric to contain only one electron, but rather that in the formulation of a breakdown criterion we need only to examine the average behaviour of one electron in the applied field. In fact, it will be seen that the presence of other electrons is implied in the formulation of the criterion.

The high energy breakdown criterion for ionic crystals was formulated by Fröhlich (1937). It is a theory of intrinsic breakdown based on the idea that indefinitely large multiplication of electrons in conduction levels will destroy the crystal. This multiplication of conduction electrons is considered to occur if the field strength is sufficiently high that ionization from the valence band can no longer be balanced by recombination to it.

Fig. 2



The average rate of energy gain from an applied field $A\ (F, E)$ for various field strengths, and the average rate of energy loss to the lattice $B\ (E)$.

This critical field strength is determined from

$$A(\mathscr{I}, F^*) = B(\mathscr{I}) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

where $\mathscr{I}$ is the ionization energy. In order to see that this equation does in fact represent the critical situation described we have plotted eqns. (22) and (30) in fig. 2 showing the average rate of energy gain from an applied field and energy loss to the lattice as a function of the energy of the electron. Consider three values of the field strength in $A(F, E)$ one of which, F^*, is such that the curves $B(E)$ and $A(F^*, E)$ intersect at $E = \mathscr{I}$, the ionization energy. Let the fields F_1 and F_2 be lower and higher than F^*. Then in an applied field F_1 an electron must have an energy E_1 if it is to be, on the average, accelerated by the field. If the applied field is F_2 then, on the average, electrons of energy greater than E_2 will be accelerated by the field. The breakdown field strength is then conceived as that field in which ionization processes cannot be balanced by the

inverse processes of recombination. It is easily seen that for a field greater than F^* the high energy electron resulting from a recombining collision is not, on the average, able to lose its energy to the lattice. Since it undergoes an acceleration it will cause a further ionizing collision, thus nullifying the effect of the recombining collision as a transfer mechanism. The high energy criterion thus gives a field strength above which recombining collisions are unable to effect a net removal of conduction electrons. The field F^* is found from (22), (13) and (30) to be

$$F^* = \frac{(6\pi^2)^{1/3}}{4\sqrt{2}} \frac{e}{\mathscr{I} a} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right) \left(\frac{m\omega_l^3 \hbar \epsilon_0 \log \gamma}{\epsilon_\infty}\right)^{1/2} \left\{1 + \frac{2}{\exp(\hbar\omega_l/kT) - 1}\right\}^{1/2} \quad (40)$$

where

$$\gamma = \frac{(6\pi^2)^{1/2} \mathscr{I}^{1/2}}{(m\epsilon_0/\epsilon_\infty)^{1/2} a \omega_l}.$$

This method for determining the critical field strength has been called the high energy criterion since the determining eqn. (38) deals with an electron of ionizing energy. Applied to the alkali halides it gives a critical field strength of order of 10^5 volts per cm and increasing with increasing temperature. The original Fröhlich theory gave critical field strengths of order 10^6 volts per cm and varying with temperature in the same way as (40). The reason for this big discrepancy lies in the incorrect application of the continuum method to determine the interaction energy to be used in calculating (22). Using the point charge method for calculating the interaction leads to the original formula of Fröhlich

$$F^* = \frac{2^{4/3}\pi^2}{4} \frac{e^3}{a^4 \mathscr{I} M} \left(\frac{m\hbar \log \gamma}{\omega_l}\right)^{1/2} \left\{1 + \frac{2}{\exp(\hbar\omega_l/kT) - 1}\right\}^{1/2} . \quad (40a)$$

which should always be used in practical application.

3.2 *The Low Energy Criterion for Ionic Crystals*

A criterion for intrinsic breakdown put forward by von Hippel (1935) and developed quantitatively at some length by Callen (1949) also considers the average behaviour of one electron. The critical field strength is considered to be that which is capable, on the average, of accelerating every conduction electron and is determined from

$$A(E', F^*) = B(E') \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

where E' is that value of the electron energy for which $B(E)$ is a maximum. This energy is only of the order of several phonon energies so that the method has become known as the low energy breakdown criterion. Using (13), (23) and (31) the critical field strength is given by

$$F^* = \frac{1}{2} \frac{em}{(E')^{1/2}} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right) \left(\frac{\omega_l^3 \epsilon_0^3 \log \gamma'}{\hbar \epsilon_\infty^3}\right)^{1/2} \times \left[1 + \frac{2}{\exp\{\hbar\omega_l(\epsilon_0/\epsilon_\infty)^{1/2}/k_0 T\} - 1}\right]^{1/2} . \quad (42)$$

where

$$\gamma' = \frac{4E'}{\hbar\omega_l} \left(\frac{\epsilon_\infty}{\epsilon_0}\right)^{1/2}.$$

Due to a different form for the interaction energy and a simplified treatment of the scattering geometry this result is not exactly as stated by Callen (1949). Applied to the alkali halides it gives a critical field strength of the same order as Callen and about twice that given by the high energy criterion.

3.3. *The Field Emission Breakdown Theory*

Field emission of electrons from the valence band to the conduction band in relation to breakdown was investigated by Zener (1934) who took as a breakdown criterion that the conduction electrons should exceed some given number. The theory was modified and developed by Franz (1939, 1956) who presented it as a theory of impulse thermal breakdown.

Franz has given the probability of field emission from the valance band to the conduction band by the Fowler–Nordheim type formula

$$P_{v\to c} = AF^{10/3} \exp\left\{ -\frac{\pi(2m)^{1/2}}{4\hbar eF} \mathscr{I}^{3/2} \right\} \quad . \quad . \quad . \quad . \quad (43)$$

where $A \sim 10^{-7}\,\mathrm{sec}^{-1}\,\mathrm{volt}^{-10/3}\,\mathrm{cm}^{10/3}$. If a field F is applied for a time t the conductivity will be

$$\sigma = P_{v\to c} N_v e\mu t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

where μ is the average mobility, and N_v is the number of valence electrons. Substituting (44) in (5) and integrating with respect to time from the time of the initial temperature T_i to the time of attainment of the critical temperature T_m we obtain

$$AN_v e\mu F^{16/3} \frac{t^2}{2} \exp\left\{ -\frac{\pi(2m)^{1/2}}{4\hbar eF} \mathscr{I}^{3/2} \right\} = C_v(T_m - T_i). \quad . \quad (45)$$

This equation can be solved for F by successive approximation. It should be emphasized that this field strength is a function of the time t for which it is applied. The field strengths obtained from this formula for insulators are several orders of magnitude higher than the breakdown strengths observed. Only for semiconductors (for which the internal ionization energy is low) does it appear possible to obtain reasonably low field strengths.

§ 4. Theories Involving the Behaviour of an Electron Distribution

4.1. *General Remarks on the Calculation of the Conduction Electron Distribution*

In general the distribution of conduction electrons in a strong field will be affected by :

(i) Acceleration by the applied field.

(ii) Collisions with other conduction electrons.

(iii) Collisions with lattice vibrations.

(iv) Ionization of electrons from or recombination to the valence level or trapping centres. In the case of ionization the required energy may

be supplied by another conduction electron, lattice vibrations, or the applied field. In the case of recombination the energy liberated may be given to another conduction electron, lattice vibrations, or alternatively emitted as light.

It is then possible to set up a formal Boltzmann equation with terms representing all of the above processes (cf. Heller 1951, Franz 1956). Because of the complexity of the equation, and the lack of knowledge of the probability of occurrence of many of these processes a solution has not been given.

In practice some approximate techniques must be developed. The first simplification arises by noting that ionization or recombination processes (of the electron–electron collision type) affect only those electrons of very low or very high energy. In between there exists a range in which the electron has too little energy to ionize, and yet sufficient to make a recombination improbable for it compared with those near the bottom of the conduction levels.

Recognizing this, there have been two main methods of attack on the problem. The first is valid for low densities of conduction electrons and ignores the effect of electron–electron collisions. In this case the distribution function in the intermediate range of energies is determined by the effect of the applied field and electron–lattice collisions. This treatment was given by Fröhlich (1947) who also showed that it was impossible to normalize the distribution function obtained unless the ionizations and recombinations were included. Attempts to do this with somewhat arbitrary descriptions of these processes have been made by Heller (1951), Franz (1952, 1956) and Veelken (1955).

The second treatment is valid for high densities of conduction electrons. It is assumed that electron–electron collisions are of such importance that they determine the form of the distribution function, and in fact cause it to be Maxwellian but with a temperature different from the lattice temperature (cf. Fröhlich 1947, 1952, O'Dwyer 1954, Fröhlich and Paranjape 1956, Stratton 1957). Under the influence of an applied field it is then necessary to consider the electron–lattice collisions in order to stabilize the assumed distribution in momentum space. The rate of energy transfer to the lattice is also determined by these collisions, or alternatively by transitions of electrons to isolated levels involving a net emission of lattice quanta. Finally, the temperature of the assumed distribution is calculated by equating the energy gain from the field to the energy loss to the lattice. Depending on the physical circumstances different assumptions have been made concerning the relative importance of the various processes to this calculation.

It is clear that the various methods for calculating the distribution function of the conduction electrons go hand in hand with a particular instability which they exhibit in a sufficiently high field. Thus, if it is being assumed that some mechanism of energy transfer to the lattice is predominant, then the calculation being made determines whether that

mechanism can be effective in leading to a steady state under the action of a field.

4.2. *The Low Density Approximation*

4.2.1. *The differential equation for the distribution function*

A full derivation of the differential equation for the electron distribution function in ionic crystals in the low density approximation was given by Fröhlich (1947), who considered temperatures and energies such that

$$kT \gg \hbar\omega, \quad \mathscr{E} > E \gg \hbar\omega. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

For a simple and less strict derivation of Fröhlich's equation it is convenient to introduce immediately a density function which can be written in a series of Legendre functions

$$\rho(E) = \rho_0(E) + \rho_1(E)\cos\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

The relation between ρ_0 and ρ_1 is easily derived on the assumption that the effect of the field inasmuch as it causes asymmetry† is simply to shift concentric shells in momentum space a distance $eF\tau(E)$ where the electrons in the shell are of energy $E = p^2/2m$. Then the asymmetrical part of the new density function will be

$$\rho_1 \cos\theta = -\frac{\partial \rho_0}{\partial E}\Delta E(E)\cos\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

where $\Delta E(E)$ is the shift in energy space corresponding to the shift $eF\tau(E)$ in momentum space. It is given by

$$\Delta E(E) = [\{p + eF\tau(E)\}^2 - p^2]/2m \simeq (2E/m)^{1/2} eF\tau(E). \quad . \quad . \quad (49)$$

Hence using (48) and (49) in (47) we have

$$\rho(E) = \rho_0(E) - \frac{\partial \rho_0}{\partial E} veF\tau(E)\cos\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

which result is of general validity provided τ is a function only of E.

The required differential equation is then obtained by relating the symmetrical part of the density function, $\rho_0(E)$, to the current of electrons in energy space $j(E)$. In view of our assumptions about electron scattering, we have for the special case under discussion

$$j(E) = [j(E)]_{\text{Field}} + [j(E)]_{\text{Coll.}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

since no other mechanism for altering the distribution is being considered.

Examining first that part due to the field we have

$$\begin{aligned}\left[\frac{\partial \rho}{\partial t}\right]_{\text{Field}} &= \frac{d\rho}{\partial E}\left[\frac{\partial E}{\partial t}\right]_{\text{Field}} \\ &= \left(\frac{2E}{m}\right)^{1/2} eF\cos\theta\left\{-\frac{\partial^2 \rho_0}{\partial E^2}\left(\frac{2E}{m}\right)^{1/2} eF\tau(E)\cos\theta\right. \\ &\quad \left. + \frac{\partial \rho_0}{\partial E} + \frac{\partial \rho_0}{\partial E}\left(\frac{2}{Em}\right)^{1/2} eF\tau(E)\cos\theta\right\} \quad . \quad . \quad . \quad (52)\end{aligned}$$

† It will be shown that the principal effect of the field under these conditions is to change the symmetrical part of the density function.

in which we have used $\tau(E) \propto E^{3/2}$ from (22). Averaging over all directions in (52) we have

$$\left[\frac{\partial \rho}{\partial t}\right]_{\text{Field}} = \frac{e^2F^2}{3m}\left\{2\tau\frac{\partial \rho_0}{\partial E} - 2E\tau\frac{\partial^2 \rho_0}{\partial E^2}\right\}$$
$$\simeq \frac{\partial}{\partial E}\left\{\frac{e^2F^2\tau}{3m}\left(\rho_0 - 2E\frac{\partial \rho_0}{\partial E}\right)\right\}$$

which is of the form of the diffusion equation $\partial\rho/\partial t = \partial j/\partial E$ in energy space, so that

$$[j(E)]_{\text{Field}} \simeq \frac{e^2F^2\tau}{3m}\left(\rho_0 - 2E\frac{\partial \rho_0}{\partial E}\right). \quad . \quad . \quad . \quad . \quad (53)$$

The current of electrons in energy space caused by the collisions is given by

$$[j(E)]_{\text{Coll.}} = \int_{E-\hbar\omega}^{E} \sum_{\mathbf{w}} \rho_0 P_{\mathbf{w}}{}^a \, dE$$
$$- \int_{E}^{E-\hbar\omega} \sum_{\mathbf{w}} \rho_0 P_{\mathbf{w}}{}^e \, dE \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

where the transition probabilities are given by (16) and (17) and the summations cover all possible collisions. In view of the inequalities (46) we may expand (54) as

$$[j(E)]_{\text{Coll.}} = -\sum_{\mathbf{w}} \hbar\omega (P_{\mathbf{w}}{}^e - P_{\mathbf{w}}{}^a)\rho_0$$
$$- \sum_{\mathbf{w}} \tfrac{1}{2}(\hbar\omega)^2 (P_{\mathbf{w}}{}^e + P_{\mathbf{w}}{}^a)\frac{\partial \rho_0}{\partial E}$$
$$- \rho_0 \frac{\partial}{\partial E}\left[\sum_{\mathbf{w}} \tfrac{1}{2}(\hbar\omega)^2 (P_{\mathbf{w}}{}^e + P_{\mathbf{w}}{}^a)\right]. \quad . \quad (55)$$

Recalling the definition of $B(E)$ from (29) we find with the use of (16) and (17)

$$\sum_{\mathbf{w}} \tfrac{1}{2}(\hbar\omega)^2 (P_{\mathbf{w}}{}^e + P_{\mathbf{w}}{}^a) = \tfrac{1}{2}\hbar\omega B(E)\{1 + 2n_{\mathbf{w}}\}$$
$$\simeq kTB(E).$$

Also from (30) and (31) we have

$$dB/dE \simeq -B/2E$$

so that (55) becomes

$$[j(E)]_{\text{Coll.}} = -B(E)\left[kT\frac{\partial \rho_0}{\partial E} + \left(1 - \frac{kT}{2E}\right)\rho_0\right]. \quad . \quad . \quad (56)$$

From (51), (53) and (56) we then have finally

$$j(E) = -\left[B(E)\left\{1 - \frac{kT}{2E}\right\} - \frac{e^2F^2}{3m}\tau(E)\right]\rho_0(E)$$
$$- \left[kTB(E) + 2E\frac{e^2F^2}{3m}\tau(E)\right]\frac{\partial \rho_0(E)}{\partial E}. \quad . \quad . \quad . \quad (57)$$

If the boundary condition is put $j(E) = 0$ then the density function will be given by

$$\rho_0(E) = \text{Const.} \times \exp\left[-\int_0^E \frac{B(E')\left\{1 - \frac{kT}{2E'}\right\} - \frac{e^2F^2}{3m}\tau(E')}{kTB(E') + 2E'\frac{e^2F^2}{3m}\tau(E')}\,dE'\right]. \quad (58)$$

For small energies the terms of the integrand in $B(E)$ are the largest and the result approximates to a Maxwellian distribution. On the other hand $\tau(E)/B(E)$ increases indefinitely for large energies and $\rho_0(E)$ approaches a finite non-zero value for any field F. This means that the density function cannot be normalized, and processes other than electron–phonon collisions must be considered to achieve a stationary state. A similar result has been given by Stratton (1957) for non-polar crystals.

4.2.2 *The collision ionization breakdown theory*

Since the distribution function cannot be normalized for the low density approximation it is natural to incorporate ionization and recombination processes in the formalism of the preceding section. This has been done by Franz (1952, 1956) and Veelken (1955) with a view to calculating an impulse thermal breakdown strength. The following are the steps in their method of calculation :

(i) For energies of the order of the ionization energy the current in energy space is put equal to the total ionization rate so that (44) can be used to relate a mean collision ionization probability w_j per electron to the applied field strength†. This relation is given in the form of an integral equation.

(ii) A given sample of dielectric of thickness d is considered between two electrodes. Then in a field F the rate of variation of the number n of conduction electrons is represented by

$$\frac{dn}{dt} = w_j n - \frac{\mu F}{d} n \qquad . \; . \; . \; . \; . \; . \; . \; . \; (59)$$

where μ is the mean mobility. Hence at a time t after the application of a field

$$n = n_0 \exp\left[\left\{w_j - \frac{\mu F}{d}\right\} t\right]$$

and the Joule heat generated by the flow of these electrons can be calculated. Thus for a sample of given thickness, density and specific heat in which the electron mobility is known the mean collision ionization probability is related to the field strength and the time required to melt the sample.

(iii) It is now assumed that voltage pulses will be applied to the sample, and the pulse time is put equal to the time required to melt the sample. The mean collision ionization probability can then be eliminated from the two equations leaving a relation for F in terms of other known constants of the sample. This is referred to as the collision ionization breakdown strength, and physically speaking it is that field which when applied as an impulse for a given time melts the sample in the same time

† In carrying out this step recombination processes are ignored. It has been noted above that a given ionization rate does not necessarily imply a current of electrons in energy space, since in a steady state each high energy electron lost by an ionizing collision is replaced by one from a recombining collision.

by means of the Joule heat generated by conduction electrons produced by collision ionization.

It should be noted that this is a theory of impulse thermal breakdown and that the breakdown field strength is therefore a function of the time of application of the voltage. Moreover, the theory contains two violations against the condition for a quasi steady state. In the first place electrons are considered as being raised from the valence band without consideration of any mechanism of recombination, and in the second electrons are being removed from the conduction level at the anode without any replacement in the conduction level at the cathode†. As we shall discuss below it seems possible that such a theory may apply in the breakdown of certain materials at high temperatures.

4.3. *The High Density Approximation*

The high density approximation is applicable when the conduction electrons exchange energy amongst themselves much more rapidly than with the lattice. Under these circumstances the energy distribution of the conduction electrons is Maxwellian at some electronic temperature T different from the lattice temperature T_0. This treatment was first introduced by Fröhlich (1947), and the following method for estimating the density of conduction electrons required in order that such an assumption be valid was given by Fröhlich and Paranjape (1956).

Consider electrons with thermal energies at some temperature greater than the lattice temperature but of the same order. Then the most energetic phonon that can interact with a given electron in this relatively low energy range will have a wave number twice the wave number of the electron, and it is reasonable to suppose that the average phonon interacting with an electron of momentum $p = \hbar k$ will have an energy of order of one half the maximum. The average energy exchange per collision for an electron of wave number k is then of the order of $\hbar sk = (Ems^2)^{1/2}$. The ratio of the probabilities of emission and absorption of a lattice quantum is $(1+n_w)/n_w$ so that the net energy loss is $1/(1+2n_w)$ so that the net energy loss is $1/(1+2n_w)$ times the energy transferred provided that $E > k_0 T_0$. If then $T(E, T_0)$ is the average time between electron–phonon collisions, then the order of magnitude of the rate of energy transfer from electrons of energy E to the lattice is

$$\left(\frac{dE}{dt}\right)_L \simeq -\frac{(Ems^2)^{1/2}}{\tau(E,T_0)}\frac{1}{1+2n_w}$$
$$\simeq -\frac{E}{k_0 T_0}\frac{ms^2}{\tau(E,T_0)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (60)$$

† The idea that electrons from the cathode enter the insulator in the valence band and are then ionized to conduction levels can only be invoked if the mobilities of conduction electrons and valence holes are equal. Otherwise a space-charge will be formed sufficient to cause a local field to procure the electrons from the cathode for the conduction levels by field emission (cf. von Hippel and Alger 1949, O'Dwyer 1954, and Rose 1955).

since $n_w \simeq k_0 T_0/\hbar\omega \simeq k_0 T_0/(Ems^2)^{1/2}$ for the average mode interacting with an electron of energy E. On the other hand the rate of loss of energy by such an electron due to collisions with other electrons is

$$\left(\frac{dE}{dt}\right)_e \simeq \frac{4\pi n e^{*4}}{p} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

(cf. Pines 1953) where e^* is an effective electronic charge differing from e by the factor depending on the dielectric constant, and n is the conduction electron density. Thus, in so far as energy transfer is concerned, we find from (60) and (61) that electron–phonon collisions and electron–electron collisions are of equal importance at the critical density

$$n_c \simeq \frac{1}{4\pi} \frac{E^{3/2} m^{3/2} s^2}{k_0 T_0 e^{*4} \tau(E, T_0)} . \quad . \quad . \quad . \quad . \quad . \quad (62)$$

With $\tau(E, T_0)$ given by (35) we have

$$n_c \sim \frac{1}{50} \frac{C^2 m^3 E^2}{\hbar^4 e^{*4} MN} . \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)$$

Putting $E \gtrsim k_0 T_0$, $(e^*/e)^4 \sim 10^{-1}$, $C \sim 1$ ev and $MN \sim 1\ \mathrm{g\ cm^{-3}}$ we find that n_c may be as low as 10^{14} per $\mathrm{cm^3}$. This critical number of electrons increases with the square of the average electronic energy which in turn increases with the strength of the applied field. However it will appear that the electronic temperature near breakdown is not more than several orders of magnitude more than the lattice temperature in the main cases of interest. It is therefore reasonable to suppose that the assumption of an electronic temperature different from the lattice temperature will apply in these cases.

For ionic crystals the high density approximation is most easily justified for low electronic temperatures $T \ll \hbar\omega/k_0$, and hence also for low lattice temperatures. Most electrons cannot then emit a quantum, and scattering is elastic with a relaxation time given by (28) and no energy transfer. Any energy transfer can only be due to the small number of electrons whose energy is large enough to emit a quantum, or to scattering by the accoustical modes. Since both processes will give a small energy transfer the high density approximation should be valid for very low densities of conduction electrons provided that the temperature is low.

For higher temperatures in ionic crystals we have in place of (60)

$$\left(\frac{dE}{dt}\right)_L = \frac{\hbar\omega}{\tau(E, T_0)} \frac{1}{1 + 2n_w} \quad . \quad . \quad . \quad . \quad . \quad (60\,a)$$

since the energy transfer per collision is the constant $\hbar\omega$. Since (61) still holds the equation for the critical electron density at which electron–phonon collisions and electron–electron collisions are of equal importance is

$$n_c \simeq \frac{1}{4\pi} \frac{p}{e^{*4}} \frac{\hbar\omega}{\tau(E, T_0)} \frac{1}{1 + 2n_w} . \quad . \quad . \quad . \quad . \quad (62\,a)$$

Extrapolating $\tau(E, T_0)$ as given by (23) to energies lower than those for which it is valid (62 *a*) becomes

$$n_c \sim \frac{me^2\omega^2}{4\pi e^{*4}}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right). \quad . \quad . \quad . \quad . \quad . \quad (63\,a)$$

Substituting values typical for the alkali halides we find that is of the order of 10^{18} per cm^3. This value is independent of the electron temperature or the applied field strength.

Under the circumstances in which the high density approximation holds, whether in polar or non-polar crystals, the distribution function can be written

$$f(\mathbf{p}) = a \exp\{-|\mathbf{p}-\mathbf{p}_0|^2/2mk_0T\} \quad . \quad . \quad . \quad . \quad (64)$$

where $N = \int f(\mathbf{p})d^3p$. This distribution function contains two arbitrary parameters p_0 and T which are to be determined from the rate of energy and momentum exchange with the lattice. The momentum displacement p_0 will be found to be proportional to the field and subject to the inequality

$$p_0^2 \ll mk_0T. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

If $f(\mathbf{p})$ is developed on a series of spherical harmonics with the field direction as axis

$$f(\mathbf{p}) = f_0(\mathbf{p}) + f_1(\mathbf{p}) \cos\theta + \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

where θ is the angle between $\mathbf{p}$ and $\mathbf{F}$, then in view of the approximation (65)

$$f_0 \simeq a \exp(-p^2/2mk_0T) \quad . \quad . \quad . \quad . \quad . \quad (67)$$

and similarly to (48)

$$f_1 = -p_0\frac{\partial f_0}{\partial p}$$

$$\simeq \frac{pp_0}{mk_0T}f_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (68)$$

The condition $n > n_c$ justifies only the Maxwellian form of the energy distribution, whereas (68) requires that the rate of momentum exchange amongst electrons be large compared with that between electrons and lattice vibrations. Stratton (1957) has shown that for non-polar crystals this requires electron densities of order $(k_0T_0/ms^2)n_c$. At room temperature this is about $10^3 n_c$, a density which may well be exceeded near breakdown. In any case (68) is a reasonable approximation, since other possible forms for f_1 would alter the final results only by a slight numerical factor.

4.3.1. *The Breakdown of pure crystals*

Within the framework of the high density approximation a theory of intrinsic breakdown has been developed (Fröhlich 1952, Paranjape 1953, 1954, O'Dwyer 1954, and Fröhlich and Paranjape 1956) for the case of a pure crystal containing a negligible number of isolated electron energy levels. The method of Fröhlich and Paranjape is given as being the most complete exposition of this topic.

In steady state conditions which apply before intrinsic breakdown the distribution function in momentum space $f(p)$ satisfies

$$\left(\frac{\partial f}{\partial t}\right)_F + \left(\frac{\partial f}{\partial t}\right)_L + \left(\frac{\partial f}{\partial t}\right)_e = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (69)$$

where the subscripts F, L and e refer to rates of change due to the field, the interaction with the lattice and the interaction with other electrons. Since electron–electron collisions conserve energy and momentum

and

$$\left.\begin{aligned} \sum_{\mathbf{p}} \mathbf{p}\left(\frac{\partial f}{\partial t}\right)_e &= 0 \\ \sum_{\mathbf{p}} \frac{p^2}{2m}\left\{\left(\frac{\partial f}{\partial t}\right)_e\right\} &= 0. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (70)$$

It follows from (69) and (70) that

$$\sum_{\mathbf{p}} \mathbf{p}\left\{\left(\frac{\partial f}{\partial t}\right)_F + \left(\frac{\partial f}{\partial t}\right)_L\right\} = 0 \quad . \quad . \quad . \quad . \quad . \quad (71)$$

$$\sum_{\mathbf{p}} \frac{p^2}{2m}\left\{\left(\frac{\partial f}{\partial t}\right)_F + \left(\frac{\partial f}{\partial t}\right)_L\right\} = 0. \quad . \quad . \quad . \quad . \quad . \quad (72)$$

These two equations are those from which the parameters p_0 and T of the distribution function (64) can be determined.

We shall consider firstly the evaluation of the terms in (71) and (72) due to the applied field and secondly the terms due to electron–lattice interaction. If the field is in the z-direction then it changes only the z component of the momentum so that

$$\left(\frac{\partial f}{\partial t}\right)_F = \frac{\partial f}{\partial p_z}\frac{\partial p_z}{\partial t} = -eF\frac{\partial f}{\partial p_z}. \quad . \quad . \quad . \quad . \quad . \quad (73)$$

Changing the summation over the momentum of (71) and (72) into an intergration we find

$$\begin{aligned} \int \mathbf{p}\left(\frac{\partial f}{\partial t}\right)_F d^3p &= -eF\int p_z\frac{\partial f}{\partial p_z}d^3p \\ &= -eFN \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (74) \end{aligned}$$

from the normalization condition of (64). Also

$$\begin{aligned} \int \frac{p^2}{2m}\left(\frac{\partial f}{\partial t}\right)_F d^3p &= -\frac{eF}{2m}\int p^2\frac{\partial f}{\partial p_z}d^3p \\ &= -\frac{eF}{m}\int p_z f\, d^3p \\ &= -\frac{eF}{m}p_0 N. \quad . \quad . \quad . \quad . \quad . \quad . \quad (75) \end{aligned}$$

The terms due to the electron–lattice interaction are more difficult to calculate. For convenience we describe the phonon by $\mathbf{q} = \hbar\mathbf{w}$ instead

of by **w**. Then the formal equation for the rate of change of f due to phonon scattering is

$$\left(\frac{\partial f}{\partial t}\right)_L = -\sum_{\mathbf{q}}[f(\mathbf{p})P^a(\mathbf{p},\mathbf{p}+\mathbf{q}) - f(\mathbf{p}+\mathbf{q})P^e(\mathbf{p}+\mathbf{q},\mathbf{p}) + f(\mathbf{p})P^e(\mathbf{p},\mathbf{p}-\mathbf{q}) - f(\mathbf{p}+\mathbf{q})P^a(\mathbf{p}-\mathbf{q},\mathbf{p})] \quad . \quad . \quad . \quad . \quad . \quad (76)$$

where the probabilities P^a and P^e for absorption and emission are given by (16) and (17) for polar crystals and (32) and (33) for non-polar crystals. The summation over $\mathbf{q}$ can be replaced by an integration in which the delta functions of the probabilities can be suitably eliminated. If in addition the distribution function in (76) is expanded (for its various arguments) up to first order in p_0 we can write

$$\left(\frac{\partial f}{\partial t}\right)_L = g_0(p) + p_0 \cos\theta g_1(p) \quad . \quad . \quad . \quad . \quad . \quad . \quad (77)$$

in which $g_0(p)$ and $g_1(p)$ can be determined from the integral form of (76) (cf. Fröhlich and Paranjape 1956).

With the method for determining $g_0(p)$ and $g_1(p)$ now established we are in a position to determine p_0 and T. Thus from (77) we have

$$\int p_z\left(\frac{\partial f}{\partial t}\right)_L d^3p = \frac{p_0}{3}\int p g_1(p)\, d^3p, \quad . \quad . \quad . \quad . \quad . \quad (78)$$

$$\int \frac{p^2}{2m}\left(\frac{\partial f}{\partial t}\right)_L d^3p = \int \frac{p^2}{2m} g_0(p)\, d^3p. \quad . \quad . \quad . \quad . \quad (79)$$

Using (74), (75), (78) and (79) in the conditions (71) and (72) gives

$$eFN = \tfrac{1}{3} p_0 \int p g_1(p)\, d^3p \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (80)$$

$$eFNp_0 = \tfrac{1}{2}\int p^2 g_0(p)\, d^3p. \quad . \quad . \quad . \quad . \quad . \quad . \quad (81)$$

Equation (80) shows that p_0 is indeed proportional to F, while (81) can be used to calculate the electronic temperature T in terms of either F or p_0. (It will be recalled that T enters into $g_0(p)$ and $g_1(p)$ when the distribution function (64) is expanded in powers of p_0 and substituted in (76).) In order to express the equation for this electronic temperature we eliminate p_0 from (80) and (81) giving

$$e^2F^2N^2 = \tfrac{1}{6}\int p^2 g_0(p)\, d^3p \int p g_1(p)\, d^3p. \quad . \quad . \quad . \quad . \quad . \quad . \quad (82)$$

From the general formulae given by Fröhlich and Paranjape for $g_0(p)$ and $g_1(p)$ it appears that g_0 vanishes for $T=0$ and both g_0 and g_1 vanish for $T\to\infty$. The quantity on the right-hand side of (82) thus has a maximum value for a certain electronic temperature T^* which defines the

maximum value F^* of the applied field for which (82) has a solution. In fact it can be seen that (82) expresses the relation

$$A(F, T, T_0) = B(T, T_0) \quad . \quad . \quad . \quad . \quad . \quad (83)$$

(cf. (39) and (41)) where A is now the energy gain from an applied field by the electron distribution at temperature T with the lattice at temperature T_0, and B is the energy loss to the lattice by the distribution. Since we may write $A = \sigma F^2$ the right-hand side of (82) is essentially B/σ where σ is a function of T and T_0. It is easily shown that for small applied fields this energy balance equation has two solutions for T (the lower one stable and the upper unstable), and that with increasing field strength these two solutions move closer to each other until they coalesce at the critical temperature T^* and the critical field F^*. For larger fields there is no solution for the electronic temperature and a stationary state is impossible.

While the problem of finding the critical field strength is now solved at its most general level, the transition probabilities P of (76) must be known before g_0 and g_1 can be calculated in any specific case. Considering first non-polar crystals for which the transition probabilities are given by (32) and (33), Fröhlich and Paranjape have calculated g_0 and g_1 algebraically and determined the breakdown field strength from (82) by numerical methods. They deal with the two cases in which the lattice temperature is either very high or very low compared with the Debye temperature Θ and find

$$\left.\begin{aligned} F^* &\simeq 1{\cdot}25 \frac{C^2 m}{e\hbar M s}\left(\frac{ms^2}{k\Theta}\right)^{1/2}\left(\frac{T_0}{\Theta}\right)^{1/2}, \text{ if } T_0 \gg \Theta, \\ F^* &\simeq 0{\cdot}71 \frac{C^2 m}{e\hbar M s}\left(\frac{ms^2}{k\Theta}\right)^{1/2}, \text{ if } T_0 \gg \Theta. \end{aligned}\right\} \quad . \quad (84)$$

The corresponding critical temperature is given by $kT^* \sim 1$ ev. As mentioned previously it would be very desirable to eliminate the constant C from (84) by introducing another experimentally measurable quantity which also depends on C in a known manner. The electronic mobility $\mu(T_0)$ is such a measurable quantity and in fact using (38) in (84) we readily find

$$\left.\begin{aligned} F^* &\simeq 0{\cdot}30 \frac{\hbar N_a^{1/3}}{m\mu(\Theta)}\left(\frac{T_0}{\Theta}\right)^{1/2}, \text{ if } T_0 \gg \Theta \\ F^* &\simeq 0{\cdot}17 \frac{\hbar N_a^{1/3}}{m\mu(\Theta)}, \text{ if } T_0 \ll \Theta \end{aligned}\right\} \quad . \quad . \quad . \quad (84a)$$

in which $\mu(\Theta)$ is the electronic mobility at the Debye temperature in a weak field.

The application to ionic crystals using the transition probabilities (16) and (17) was also given by Fröhlich and Paranjape and a small correction

to their work has been made by Stratton (personal communication) for the high temperature case. The results are

$$\left.\begin{aligned} F^* &\simeq \frac{me\omega_l}{\hbar}\left(\frac{\epsilon_0}{\epsilon_\infty}\right)^{1/2}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon_0}\right)\left[\frac{2}{3\pi}\log\left(\frac{1{\cdot}6\,T_0}{\Theta}\right)\right]^{1/2}, \text{ if } T_0 \gg \Theta, \\ F^* &\simeq 0{\cdot}5\frac{me\omega_l}{\hbar}\left(\frac{\epsilon_0}{\epsilon_\infty}\right)^{1/2}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon_0}\right), \text{ if } T_0 \ll \Theta. \end{aligned}\right\} \quad (85)$$

In this case the critical temperatures are found to be given by $k_0 T^* \sim$ several $\hbar\omega$, so that ω will be given by (27). The low temperature case of (85) can also be written in terms of the mobility with the help of (28) giving

$$F^* \simeq 0{\cdot}5\left(\frac{\hbar\omega_l}{2m}\sqrt{\frac{\epsilon_0}{\epsilon_\infty}}\right)^{1/2}\frac{1}{\mu(T_0)n_w(T_0)} \quad . \quad . \quad . \quad . \quad (85\,a)$$

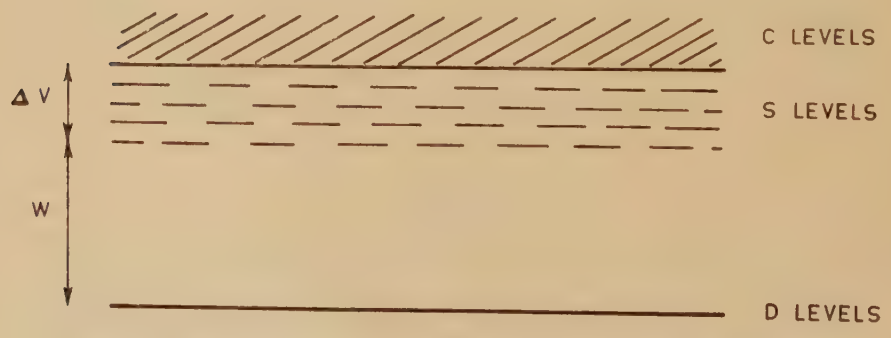
if $T_0 \ll \Theta$ and $T_0 \ll \hbar\omega/k_0$.

The equations for the critical field strength in non-polar crystals utilize the results for perturbation theory in a region in which it should be accurate. It is therefore expected that (84 *a*) should give an accurate value for the breakdown field strength of a non-polar insulator which breaks down according to the process described. For polar crystals, however, perturbation theory is extrapolated outside its known region of validity by these calculations. The only justification would seem to be that a similar extrapolation gives reasonable results in the calculation of electronic mobility.

4.3.2. *The breakdown of crystals containing imperfections*

A theory of intrinsic breakdown has been given by Fröhlich (1947) for the case in which a relatively large number of crystal imperfections give rise to isolated electron energy levels below the conduction band. It is assumed that these energy levels are spread over an energy interval ΔV immediately below the conduction band, and are separated from lower levels by a much larger energy W. The energy level diagram for electrons is then as shown in fig. 3. It is further assumed that electrons

Fig. 3



Energy level diagram for a model of an amorphous dielectric.

in C and S levels together are in thermal equilibrium at a temperature T different from the lattice temperature T_0. The more complicated analysis of the preceding sections will be unnecessary for our present purpose, the assumption of a Maxwellian energy distribution of energy being sufficient. The number of C and S electrons required to justify this assumption will no longer be correctly given by the calculations above, but it can be assumed that it would be of a similar order of magnitude. If N_C, N_S and N_D are the numbers of electrons in conduction levels, shallow traps and deep traps respectively then it is easily shown that

$$\left.\begin{aligned} &N_C \ll N_S \ll N_D \\ \text{provided that} \quad &\\ &k_0 T \ll \Delta V \ll W. \end{aligned}\right\} \quad \ldots\ldots \quad (86)$$

In fact we have

$$N_C = N_S \gamma \exp(-\Delta V/k_0 T) \quad \ldots\ldots \quad (87)$$

where γ is the ratio of the density of energy levels in a range near the bottom of the C levels, to a corresponding quantity near the bottom of the S levels.

Since only the conduction electrons are free to move through the lattice then only they will receive energy from the field, while in view of the inequality (86) the electrons in the shallow traps will be mainly responsible for transferring energy to the lattice by the emission of single lattice quanta. Thus the energy balance eqn. (83) applies with this understanding and we have

$$A(F, T, T_0) = N_C \frac{e^2 \tau_1}{m} F^2 \quad \ldots\ldots \quad (88)$$

where τ_1 is a relaxation time for conduction processes, and

$$B(T, T_0) = N_S \frac{\hbar\omega}{\tau_2} \frac{\exp(\hbar\omega/k_0 T_0 - \hbar\omega/k_0 T) - 1}{\exp(\hbar\omega/k_0 T_0) - 1} \quad \ldots \quad (89)$$

where $1/\tau_2$ is an average value for the transition probability connected with the absorption or emission of a phonon. The exponential factors in (89) take account of the fact that a net energy transfer occurs only if the electron and lattice temperatures are different. It is easily derived by considering the difference between the total number of emission and absorption processes (cf. Fröhlich 1947). Using (87), (88) and (89) the energy balance equation becomes

$$DF^2 \exp(-\Delta V/k_0 T) = \exp(\hbar\omega/k_0 T_0 - \hbar\omega/k_0 T) - 1 \quad . \quad (90)$$

where

$$D = \frac{e^2 \tau_1 \tau_2}{m \hbar\omega} \gamma \{\exp(\hbar\omega/k_0 T_0) - 1\}.$$

Equation (90) may be solved for a critical field strength and critical temperature in exactly the same way as (82). In this case the operation

can be performed by algebraic approximation and we find

$$\left.\begin{aligned} \frac{1}{k_0 T_0} - \frac{1}{k_0 T^*} &\simeq \frac{1}{\Delta V} \\ F^* &\simeq C \exp\left(\Delta V / 2 k_0 T_0\right) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (91)$$

$$\text{where } C^2 \simeq \hbar\omega / 2{\cdot}71 D \Delta V.$$

The amorphous solid theory of Fröhlich does not give an absolute value for the breakdown strength, but only its temperature dependence. Using the same ideas a calculation has been given (O'Dwyer 1957) to determine the absolute value of the breakdown strength for a somewhat more special model applying only to polar crystals. The imperfections, n_D per unit volume, are considered as giving rise to a single isolated electron level $p_0\hbar\omega$ below the conduction band. As previously the conduction electrons are responsible for the energy gain from the field, while the loss to the lattice is now accounted for by transitions to the isolated level involving the simultaneous emission of many phonons.

Calculating the conduction relaxation time for a Maxwell distribution of electrons at temperature we find for $T \gg T_0$

$$A(F, T, T^0) = C_A \left(\frac{k_0 T}{\hbar\omega}\right)^{1/4} \frac{T}{T_0} F^2 \exp\left(-p_0 \hbar\omega / 2 k_0 T\right) \quad . \quad . \quad (92)$$

where C_A is independent of T. Using the method of Huang and Rhys (1950) for the multiphonon transitions in polar crystals the energy loss to the lattice is expressed as

$$B(T, T_0) = C_B \frac{k_0 T_0}{\hbar\omega} \left(\frac{S k_0 T_0}{\hbar\omega}\right)^{1/4} \exp\left(-S \frac{\hbar\omega}{k_0 T} + S \frac{\hbar\omega}{4 k_0 T_0}\right) \quad . \quad . \quad (93)$$

where C_B is also independent of T, and S is an interaction constant related to the equilibrium potential energy of the defect in the lattice. The solution of (82) for a critical temperature and field strength proceeds again as above to give

$$1/k_0 T^* = 5/2(2S - p_0)\hbar\omega \quad . \quad . \quad . \quad . \quad . \quad . \quad (94)$$

$$F^* = C f(S, p_0, T_0) \quad . \quad . \quad . \quad . \quad . \quad . \quad (95)$$

where

$$C \simeq 60 \frac{\omega(\hbar\omega)^{1/8} m^{9/8} (1/\epsilon_\infty - 1/\epsilon_0)}{\hbar^{5/4} a^{3/2}}$$

and

$$f(S, p_0, T_0) = \frac{|\langle 0|e\mathbf{x}|\mu\rangle| n_D^{1/4} k_0 T_0}{p_0 \hbar\omega} \left(S \frac{k_0 T_0}{\hbar\omega}\right)^{1/8} \times (2S - p_0)^{5/8} \exp\left(\frac{S\hbar\omega}{8 k_0 T_0}\right),$$

C is a constant of the pure substance, and $f(S, p_0, T_0)$ is determined by the electron trap. The quantity $|\langle 0|e\mathbf{x}|\mu\rangle|$ is the matrix element of the electric moment between the trapped state and the lowest state in the

conduction band. The evaluation of this critical field strength then depends on a knowledge of p_0, S and $\langle 0 \,|\, e\mathbf{x} \,|\, \mu \rangle$ for the particular trapping state being considered. These parameters can easily be evaluated for the case in which the trapping centres are interstitial positive ions or vacant negative ion sites using the methods given by Simpson (1949). The variation of breakdown strength with temperature is, however, a result of conflicting factors in (95) and it is not possible to say in general what type of characteristic is to be expected.

§ 5. Comparison with Experimental Work

5.1. *General Difficulties of Interpretation*

The theories of breakdown given above, whether intrinsic or impulse thermal, calculate a critical field strength for a homogeneous isotropic material at a given temperature and on the assumption that the current flow is purely electronic. In practice this situation is never achieved.

In the first place all experiments measure a breakdown voltage across a certain specimen distance and not an actual field strength. The maximum field strength in the specimen may differ greatly from this average field strength for various reasons. The most obvious source of difficulty in this respect comes from geometrical irregularities either of the applied electrodes or of the surfaces of the dielectric which they touch. Similar effects would be caused by chemical contamination of either electrode, particularly contamination of the cathode since electrons must be supplied from it to continue the electronic current in the dielectric. However, even in an experiment in which neither geometrical irregularities nor chemical contamination is present to any appreciable degree, the field strength may not be uniform in the specimen. This is so since current flow is not symmetrical to the extent that the process by which the cathode supplies electrons is different from the process by which the anode accepts them. In general a space-charge may be set up in a dielectric such as to regulate correctly the supply of electrons from the cathode. This causes non uniformity of the field strength in the dielectric (cf. O'Dwyer 1954, Rose 1955 and Lampert 1956).

Another source of error in experimental work is the temperature measurement. For practical reasons the temperature measured is always that of one of the electrodes as near as practicable to the anticipated breakdown path. This may not be the temperature of the dielectric at breakdown for two possible reasons—firstly in impulse testing there may be insufficient time for heat generated in the region of breakdown to be conducted to the electrodes; secondly, since, even if the heat is conducted to the electrodes, a large temperature rise in the breakdown region of the dielectric could produce quite a small temperature rise in the electrodes, on account of the great difference in heat capacity between the electrode system and the dielectric sample, and also the much greater area of thermal contact between the electrode and the ambient than between the dielectric sample

and the ambient. The constancy of intrinsic breakdown strength over wide ranges of conditions of voltage application implies that such an effect does not seriously interfere with these measurements. However for temperature ranges in which it is not clear whether breakdown is intrinsic or impulse thermal this effect may be a source of apparent disagreement, since different thermal conditions could give different electrical results.

Deviation from the theories also occurs in that the electric current is not purely electronic in all real insulators. Where intrinsic breakdown has been established it appears fair to neglect the effects of ionic currents on the grounds that they obviously do not lead to the breakdown. They could, however, lead to some form of systemmatic error in the temperature measurement. For a theory of impulse thermal breakdown the ionic currents cannot be ignored since they also contribute to the Joule heat generated. It seems difficult to establish a theory of impulse thermal breakdown on a quantitative basis for any substance in which there are appreciable electronic and ionic contributions to the current.

In most experimental work it has been tacitly assumed that the internal impedance of the high voltage supply is of no importance. This may not be so, especially since it will often be very high in the usual arrangement in which a high voltage capacitor is charged from a rectifier of low output. The energy stored in the capacitor is then all that is available to complete a breakdown once it has commenced, and if this energy is insufficient anomalous results may be obtained.

There are many other practical points which theories overlook among which could be mentioned stresses in the dielectric material and the possible effect of crystallographic directions. The stresses in the dielectric could either be due to the lack of annealing or to the large applied field and their effects have been experimentally investigated by Calderwood, Cooper and Wallace (1953) and Cooper and Wallace (1953) who have shown that stress in a crystal raises the breakdown strength. The effect of crystallographic direction on the breakdown strength has been investigated by Davisson (1946) and by Cooper, Grossart and Wallace (1957). They find that the directional effect is fairly small for alkali halides with some exceptions.

It is therefore with the reservations due to these points in mind that the following comparison is made with experimental work. Moreover a satisfactory interpretation has not been given for all of the observed facts of breakdown of any substance, but rather it has been possible up to the present only to explain some of the more general and best established of the experimental results.

5.2. *Polar Crystals*

5.2.1. *The alkali halides*

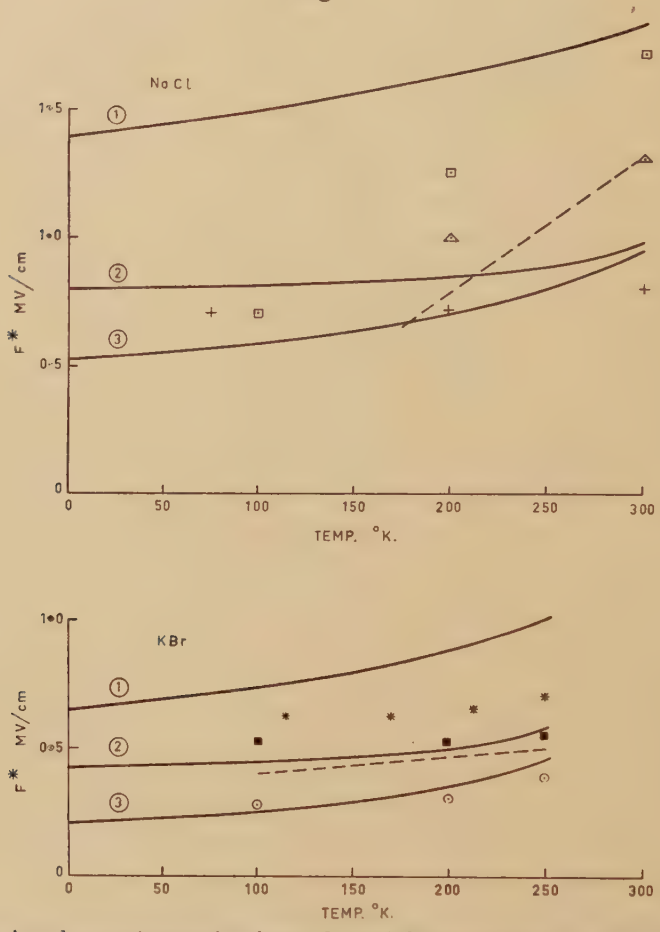
There are two types of experiment which have been used for comparison with the theory of dielectric breakdown of alkali halides. One has been

to select a particular substance and to measure its breakdown strength as a function of temperature, and the other has been to measure the breakdown strength of many substances at some fixed temperature.

The alkali halides which have been measured over a range of temperature are NaCl (von Hippel and Lee 1941, Alger and von Hippel 1949, Calderwood and Cooper 1953, and Cooper, Grossart and Wallace 1957); KCl (Calderwood and Cooper 1953, and Cooper, Grossart and Wallace 1957) and KBr (Beuhl and von Hippel 1939, Austen and Whitehead 1940, Alger and von Hippel 1949, and Konorova and Sorokina 1957.) Since NaCl and KBr have been measured by different groups of authors we shall confine our comments to these two substances. In the first place we shall discuss the low temperature region in which breakdown is intrinsic and for which the results are shown in fig. 4. The full lines show the temperature dependence of the theoretical values of the intrinsic breakdown strength according to the various theories, while the points and the dotted lines give the various experimental results. The von Hippel–Callen low energy criterion has been recognized as an upper limit to the theoretical breakdown strength since it requires that every electron taken independently be accelerated indefinitely by the applied field. The Fröhlich high energy criterion may be regarded as a lower limit since it requires only that recombination be not able to balance ionization with the electrons again considered as independent. This calculation neglects the possibility of some other process restoring equilibrium. If electron–electron collisions establish a Maxwellian energy distribution at some fairly low electronic temperature there will be no conduction electrons capable of ionizing so that the considerations of a high energy criterion are not then crucial. Thus from a theoretical point of view it seems more reasonable to regard the Fröhlich–Paranjape theory as the best calculation of the breakdown strength since it requires that every electron be accelerated indefinitely by the field, their interactions with each other being taken into account. The critical field strength of the high energy criterion can then be regarded as supplying a means of increasing the density of conduction electrons to a level at which the high density approximation is valid, if indeed this increase is required. Considering the experimental difficulties, and the scatter and lack of reproducibility of the results (some authors give different values for the breakdown strength of the same substance in different papers) the agreement between theory and experiment is fairly good. It is interesting to note that the results of the Fröhlich–Paranjape theory seem to apply well in the low temperature region where theoretical considerations best support the validity of the high density approximation. The form of the Fröhlich–Paranjape theory shown in fig. 4 is that of (85) and not of (85 *a*) since the latter can be used only at very low temperatures at which the mobility is known. Using the experimental result of Redfield (1953) that $\mu \sim 250$ cm^2 volt^{-1} sec^{-1} at 82°K for NaCl the breakdown strength is about 0·7 MEV cm^{-1} comparing favourably with 0·8 MEV cm^{-1} given by (85) at that temperature.

In the high temperature region for these substances the experimental evidence is very conflicting even to the value of the transition temperature. Alger and von Hippel and also Konorova and Sorokina have observed impulse thermal breakdown in KBr for pulses longer than one microsecond

Fig. 4



Theoretical and experimental values of breakdown strength for NaCl and KBr in the low temperature region.

in duration. To apply field emission theory is out of the question on account of the high internal ionization energy of this substance. A calculation of the breakdown strength on the basis of collision ionization

theory has been done by Franz (1952) and by Veelken (1955) for a voltage impulse of 10^{-3} sec duration. They find a value of 0·44 MEV cm^{-1} for the breakdown strength with negligible temperature variation. Considerable temperature variation is observed experimentally which might be expected, since ionic conduction can obviously not be neglected in an impulse thermal breakdown theory of an ionic crystal. Calderwood and Cooper, and later Cooper, Grossart and Wallace have observed intrinsic breakdown in the high temperature region for NaCl and KCl. The breakdown strength decreases with increasing temperature and application of the Fröhlich theory of the breakdown of amorphous solids to the results of Calderwood and Cooper gives a trap depth range of order of 0·1 ev which seems very low for ionic crystals. Considering the electron traps to be due to interstitial positive ions the critical field strength (95) gives a correct order of magnitude and temperature variation for these results. In the same temperature region Cooper, Grossart and Wallace find that KCl is strongly anisotropic but that NaCl is not. The temperature variation of breakdown strength in this work appears to be smaller than those previously reported. The situation in the high temperature region is thus confusing and some possible causes of this may be of interest:

(i) Impulse thermal breakdown has been reported in KBr and intrinsic breakdown in KCl and NaCl. It is improbable but not impossible that these alkali halides do in fact behave differently.

(ii) Cooper and Wallace (1954) have demonstrated the great importance of lattice deformations in measurements of breakdown strength. It is thus essential that any heat treatment of samples be the same if results are to be comparative.

(iii) Cooper, Grossart and Wallace have shown the strong anisotropy of KCl at temperatures above room temperature. It is therefore necessary to compare breakdown results in the same crystallographic directions.

(iv) In any case of suspected impulse thermal breakdown the heat capacity of the electrodes and their thermal and electrical contact with the sample could be of importance. In particular there could be differences due to electrodes of graphite or of evaporated metal.

(v) In cases of impulse thermal breakdown it may be questioned as to whether the voltage impulse generator was capable of supplying sufficient current over the range of pulse times for which it was used.

Turning now to measurements made on different alkali halides at one temperature, those often quoted are due to von Hippel (1935). There have been several attempts to form an empirical relation between these measured breakdown strengths and some other characteristic crystal parameter. For example Franz (1956) gives

$$F^* \simeq 26/a^3 \qquad (96)$$

where F^* is in MV cm^{-1} and the interionic distance a in Å. Vorob'ev

(1956) gives an approximate linear relation between the breakdown strength and the cohesive energy. His relationship is approximately

$$F^* \simeq 0{\cdot}6E_c - 3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (97)$$

where E_c is in ev per ion pair.

Such empirical relations are of limited interest for various reasons. In the first place it is not obvious that experimental values of F^* at some given temperature (which is the same for various substances) are suitable for comparison. In fact the Fröhlich–Paranjape theory suggests that measurements at the same reduced temperature T/Θ would make more suitable comparisons. Secondly (accepting the comparison at a given temperature) the results (96) and (97) can be derived from the Fröhlich–Paranjape theory to within the order of accuracy with which they represent experimental work.

In the Fröhlich–Paranjape theory the breakdown strength of alkali halides is given by (cf. (85))

$$F^* = G(T/\Theta)F_0$$

$$F_0 = \frac{me}{\hbar}\omega_t\left(\frac{\epsilon_0}{\epsilon_\infty}\right)^{1/2}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right) \quad . \quad . \quad . \quad . \quad . \quad (98)$$

where G is of the order of unity and a function of the reduced temperature only. We shall see that the characteristic field strength F_0 fits the variation represented by (96) and (97). For this purpose we shall require the formula

$$\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0} = \frac{4\pi}{\epsilon_0\epsilon_\infty}\left(\frac{\epsilon_\infty + 2}{3}\right)^2 \frac{e^{*2}}{2M\omega_t^2 a^3} \quad . \quad . \quad . \quad . \quad . \quad (99)$$

from the theory of the dielectric constant of ionic crystals (cf. Fröhlich (1949). Here M is the reduced ionic mass and e^* the effective ionic charge. Substituting (99) in (98) we find

$$F_0 = \left[\frac{4\pi me}{\hbar\epsilon_0\epsilon_\infty}\left(\frac{\epsilon_0}{\epsilon_\infty}\right)^{1/2}\left(\frac{\epsilon_\infty + 2}{3}\right)^2 \frac{e^{*2}}{2M\omega_t}\right]\frac{1}{a^3}. \quad . \quad . \quad (100)$$

The quantity in square brackets varies between 29 and 39 for all the halides of sodium and potassium in good agreement with (96).

The correlation with cohesive energy is as easily explained when it is noted that

$$F^* \simeq 0{\cdot}0025E_c^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (101)$$

fits the experimental points of fig. 5 as well as (97) does. The Madelung electrostatic energy sum is a good first approximation to the cohesive energy of the alkali halides, and is given by

$$E_c = 1{\cdot}75e^2/a$$

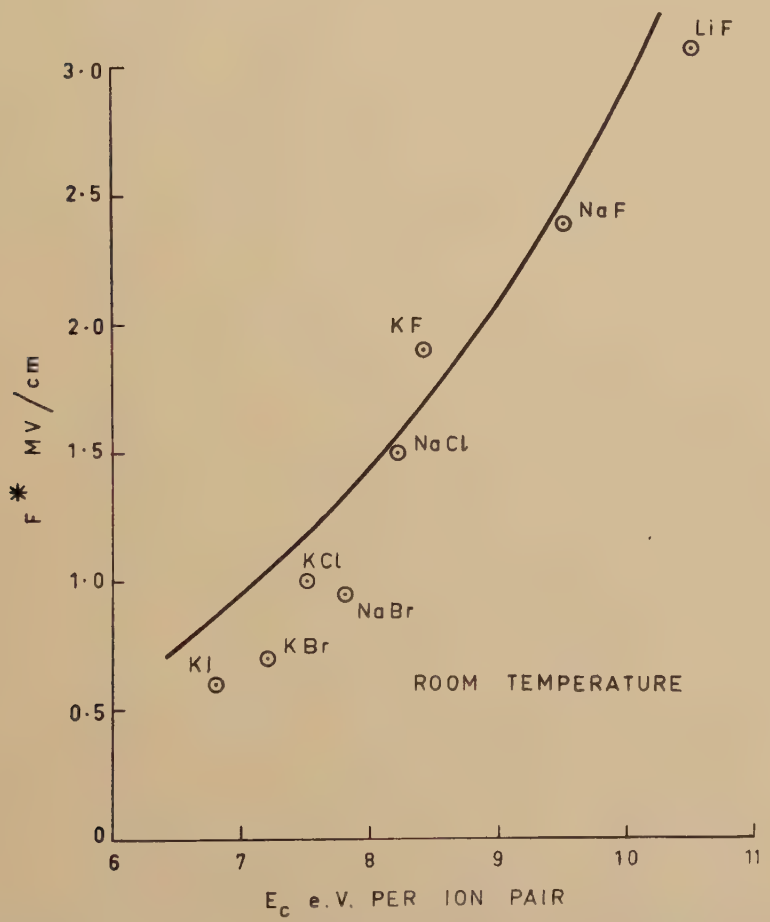
$$\simeq 25/a \text{ ev per ion pair.} \quad . \quad . \quad . \quad . \quad . \quad (102)$$

Substituting (102) in (101) gives

$$F^* \simeq 39/a^3$$

agreeing well with (100). All such empirical results do not of themselves support any mechanism of breakdown, even though it may appear tempting to speculate that breakdown is thermal when it is correlated with the cohesive energy. If a theoretical result shows good agreement with experiment, it naurally follows that all empirical relations must be at least approximately reducible to the theoretical one.

Fig. 5



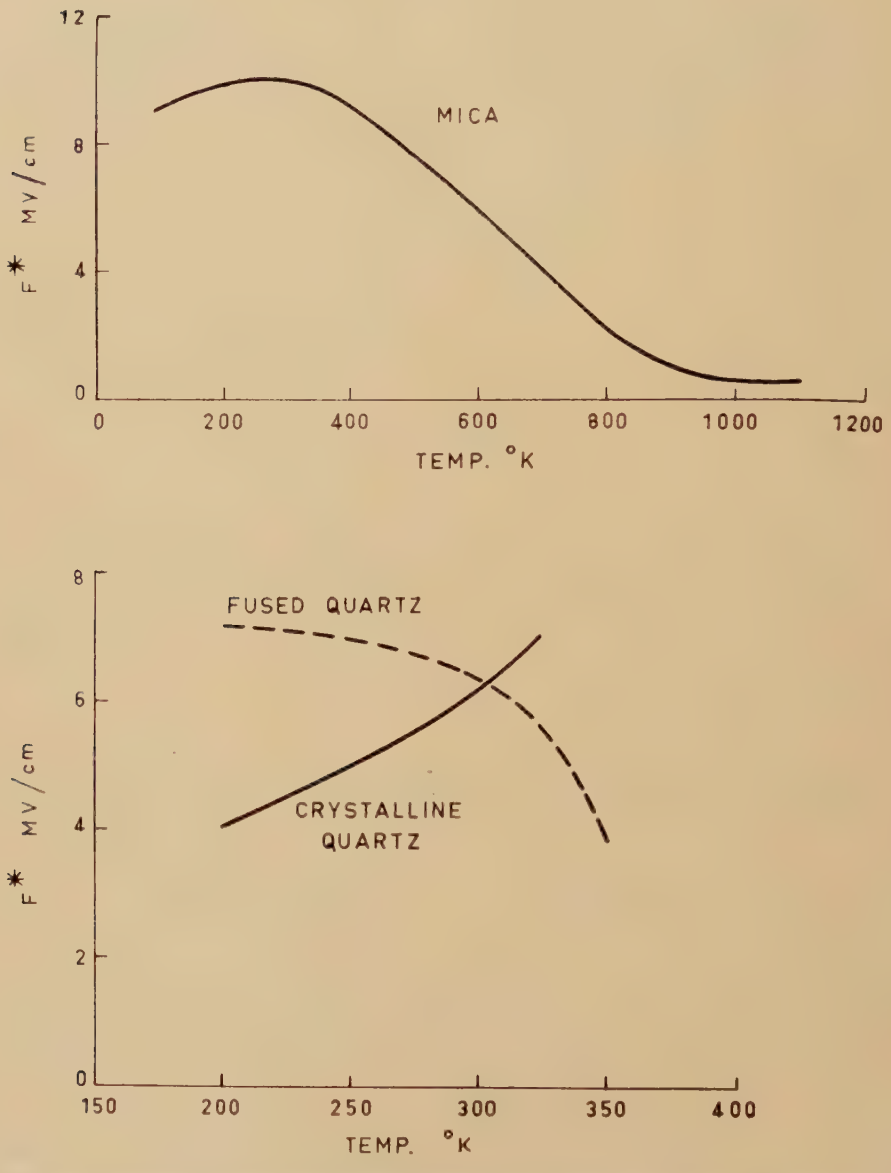
The breakdown strength of alkali halides at room temperature as a cubic function of cohesive energy.

5.2.2. *Mica and quartz*

Mica and crystalline quartz are examples of polar crystals whose breakdown strength is considerably higher than that of the alkali halides.

The variation of breakdown strength with temperature is shown for these two substances in fig. 6. It is of interest to calculate the Fröhlich–

Fig. 6



The temperature dependence of the breakdown strength of mica (Austen and Whitehead (1940)) and quartz (von Hippel and Maurer (1941)).

Paranjape characteristic field strength F_0 for these substances. There are several infra-red absorption frequencies for both quartz and mica but

the calculation of F_0 for the highest frequency would give the upper theoretical limit. Taking this absorption to have an angular frequency of $\omega_l \sim 2 \times 10^{14}$ sec^{-1} for both these substances and using the appropriate values of the dielectric constant in (98) it is found that

$$F_{0_{\text{Mica}}} \simeq 11 \text{ MV cm}^{-1}$$
$$F_{0_{\text{Quartz}}} \simeq 7 \text{ MV cm}^{-1}.$$

For both substances the temperature dependent factor G of (98) is of the order of $\frac{1}{2}$ and varies very slowly with temperature. These results are in reasonable agreement with experiment.

Two further interesting experimental points deserve mention. Many authors (Austen and Whitehead 1940, Ryu and Kawamura 1954, and Konorova 1957) have measured the breakdown strength of very thin flakes of mica, and have found an increasing breakdown strength with decreasing thickness. Considering that this thickness was evidence of avalanche breakdown Kawamura and Onuki (1953) and Kawamura, Ohkura and Kikuchi (1954) have investigated the statistical time lag of breakdown in mica, and find it to be much larger than for glass or KCl. There is thus some evidence to support the Seitz avalanche breakdown theory in this case.

Secondly, although fused quartz should be discussed as glass, we have shown the results of von Hippel and Maurer (1941) who have compared its behaviour with that of crystalline quartz. As will appear from later work on glass the breakdown of the fused quartz is probably intrinsic at the lower temperatures and impulse thermal at higher temperatures; on the other hand the breakdown of crystalline quartz appears to be intrinsic over the range of temperature shown. It is then interesting to observe that at the lower temperatures where breakdown is intrinsic for both, the fused quartz is stronger than the crystalline; also that intrinsic breakdown occurs at higher temperatures for crystalline quartz than for fused. The higher intrinsic breakdown strength of the fused quartz is easily explained on the grounds of lower electronic mobility in it; in other words compared with crystalline quartz the electron is much more greatly hindered in acquiring energy from the applied field, while its difficulty in transferring energy to the lattice is not so greatly increased. Regarding the transition temperature from intrinsic to impulse thermal breakdown it will be recalled that we have considered that this will occur when the electrical and thermal conductivities reach a certain critical relationship. In fused quartz the thermal conductivity is less than that in crystalline quartz by a factor of about ten, while the electrical conductivity is less by a factor of about three (at room temperature). It is then reasonable to expect the onset of impulse thermal breakdown at a lower temperature in fused quartz.

5.3. *Non-Polar Crystals*

The Fröhlich–Paranjape theory can be made to yield a remarkably simple numerical formula in the case of breakdown of non-polar crystals.

If the temperature is well below the Debye temperature

$$F^* \simeq 3/\mu(\Theta) \text{ mev cm}^{-1} \quad . \quad . \quad . \quad . \quad . \quad (103)$$

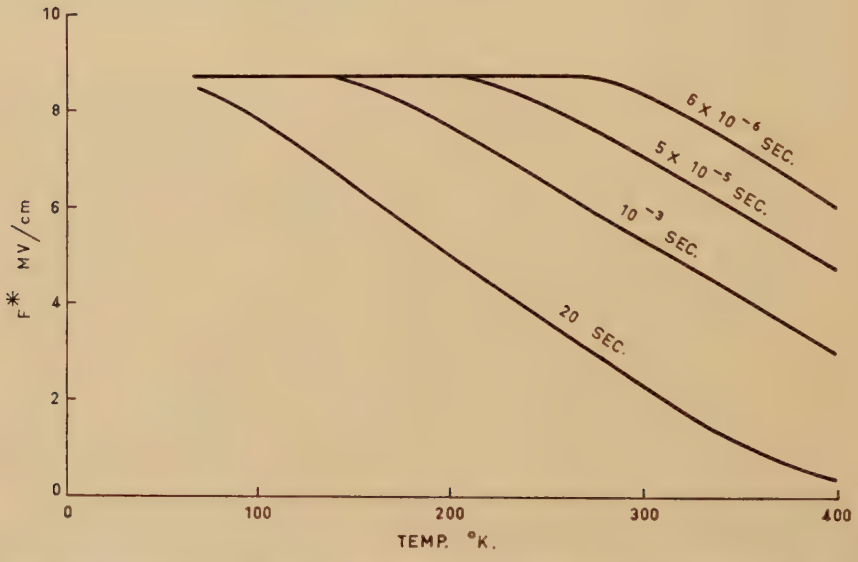
where $\mu(\Theta)$ is the mobility at the Debye temperature in $\text{cm}^2\ \text{volt}^{-1}\ \text{sec}^{-1}$. This is easily derived from (84 *a*) since we can always put $N_a^{1/3}$ of the order of 2×10^7. Equation (103) is perhaps one of the most interesting results of breakdown theory, but unfortunately there seems to be no accurate experimental work with which to compare it.

5.4. *Amorphous Solids*

5.4.1. *Glass*

Extensive series of experiments on various glasses have been performed by Keller (1948, 1952) and Vermeer (1954, 1956). In fig. 7 we show Keller's results on Thüringian glass giving the breakdown strength as a function

Fig. 7



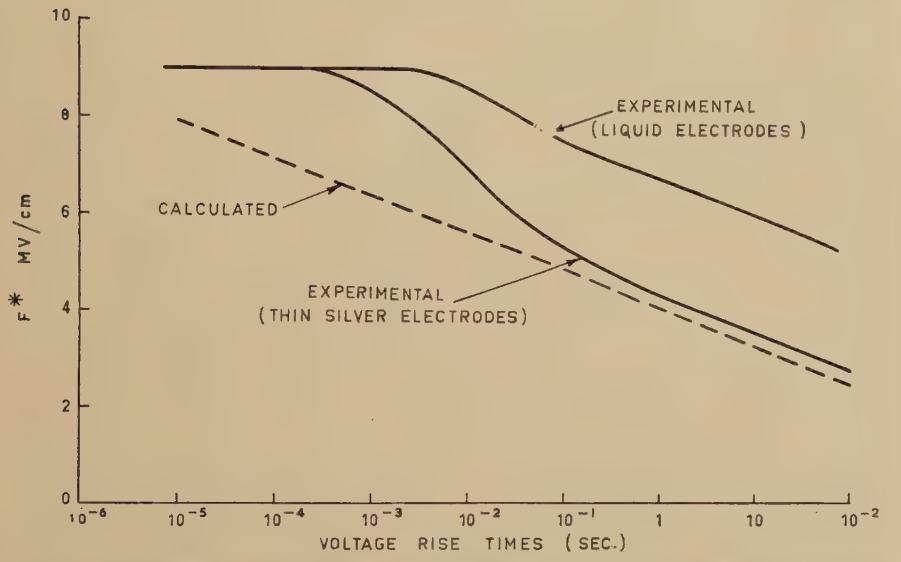
Breakdown strength of Thüringian glass as a function of temperature for various pulse rise times (Keller (1951)).

of temperature for various pulse rise times. The pulses used were such that the voltage rose linearly with time, and the dielectric sample was destroyed by the application of one pulse. In this way the effect of any possible preheating was eliminated. The diagram is typical of the situation in which breakdown is intrinsic for low temperatures or short rise times giving way to impulse thermal as the temperature or the rise time increases. Results of similar form are given by Vermeer for Pyrex, Phillips 08 and Phillips 18 glasses.

There is no satisfactory theory of the breakdown strength in the intrinsic region—it can be noted though that it should be higher than for crystalline quartz which the theory gives as of order of 5 MV cm^{-1}.

In the region of impulse thermal breakdown a somewhat phenomenological theory has been given by Vermeer (1956) who has measured the electrical conductivity as a function of temperature and field strength, and used the result in (5)—the basic equation of impulse thermal breakdown. A typical sample of Vermeer's work is shown in fig. 8 where the measured breakdown strength of Pyrex glass at 20°C with thin silver electrodes is shown as a function of the voltage rise time. The values of

Fig. 8



Experimental and calculated breakdown strengths of pyrex glass at 20°C as a function of voltage rise time (Vermeer (1956)).

σ used in the calculation are measured up to a field strength of about 1 MV cm^{-1} and extrapolated to higher field strengths. A feature of these results is the excellent agreement between measurements and calculations for the voltage rise times longer than about 10^{-1} sec. It would seem that this can be attributed to the relatively small extrapolation of the measured values of σ, while for shorter voltage rise times the higher breakdown strengths would require unreasonably large extrapolations.

It is interesting to enquire over what range of voltage rise times impulse thermal breakdown could be expected to occur. Equation (4) provides a basis for estimating this since the assumptions of impulse thermal breakdown require that the heat conduction term be very small compared with the temperature rise term. An order of magnitude expression for the equality of these terms is

$$t = \frac{C_v l^2}{K} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (104)$$

where l is a length associated with the dielectric sample, and may be taken as the thickness. Equation (104) may be regarded as giving the time of the longest pulse for which breakdown can be considered as impulse thermal; for appreciably longer pulse rise times thermal conduction processes must be considered. For glass at room temperature $K \sim 10^{-3}$ cal cm^{-1} sec^{-1} °C^{-1}, $C_v \sim 1$ cal cm^{-3} °C^{-1} and in Vermeer's work $l \sim 30 \times 10^{-4}$ cm, so that $t \sim 10^{-2}$ sec. The good agreement obtained by Vermeer for longer voltage rise times could be explained if the electrodes were poor thermal sinks, thus effectively increasing the dimension l to be used in (104). This was in fact so since the electrodes were evaporated silver films. For liquid electrodes the thermal properties of the electrodes resulted in higher breakdown strengths in the impulse thermal region as shown in fig. 8. Keller (1948) has shown that mercury electrodes produce an even greater effect.

As in all thermal breakdown the destruction of the sample was considered by Vermeer to be the result of melting due to temperature rise which occurs slowly at first and then very rapidly. It should be noted, however, that there appears to be nothing to prevent the sample from failing by an intrinsic type of breakdown, if for sufficiently short pulses the breakdown strength decreases with increasing temperature. The voltage pulse could then heat the dielectric to a temperature at which an intrinsic type mechanism occurs, and breakdown would not then occur simply as a continuation of the already initiated heated process. This point is not of great practical importance, but may be of some interest in the study of the breakdown mechanism.

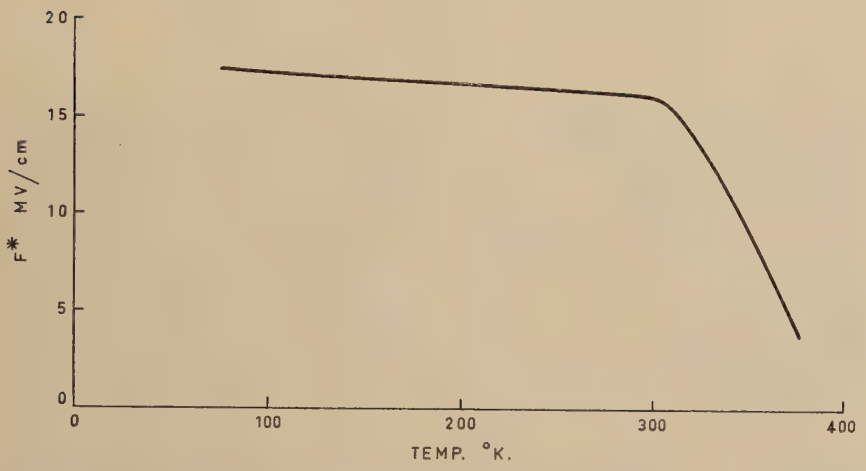
5.4.2. *Polymers*

The temperature dependence of the breakdown strength of many polymers has been measured (see for example Whitehead 1950), and as a typical example we give in fig. 9 the results of Oakes (1948) on polythene. In these experiments the rise time of the applied voltage was so long that we may consider them to be d.c. tests. It would, at first sight, seem natural to assume that a similar discussion to that given for glass would hold in the case of polythene, but an important difference is worth noting. Since the thermal conductivity is about one tenth that for glass, and the electrical conductivity less by many powers of ten it is probable that the breakdown of polythene is intrinsic over the whole of the measured range. The change in slope of the breakdown strength versus temperature curve which occurs at room temperature can be explained on the grounds that polythene is mainly crystalline below this temperature and mainly amorphous above it. Under these circumstances the Fröhlich theory of intrinsic breakdown of amorphous solids may hold above room temperature. Applying eqn. (91) to the results of fig. 9 we find that the range of depth of shallow traps would be given by $\Delta V = 0{\cdot}36$ ev.

This conception has some support from the work of Fowler (1956) who has measured the conductivity induced in polythene by ionizing

radiation. Fowler has found that the temperature variation of this induced conductivity has an equivalent activation energy of 0·4 ev which is in excellent agreement with that found from breakdown results.

Fig. 9



Temperature dependence of the breakdown strength of polythene (Oakes (1948)).

Acknowledgments

The author wishes to express his gratitude to Professor H. Fröhlich, F.R.S., of the University of Liverpool for his encouragement and advice in connection with this work. He is also indebted to Mr. F. J. Lehany, Chief of the Division of Electrotechnology, and his colleagues at the National Standards Laboratory for helpful discussions; also to Dr. R. Stratton of Metropolitan-Vickers for kindly supplying unpublished research work.

References

Alger, R. S., and von Hippel, A., 1949, *Phys. Rev.*, **76,** 127.
Austen, E. W., and Whitehead, S., 1940, *Proc. roy. Soc.* A, **176,** 33.
Buehl, R. C., and von Hippel, A., 1939, *Phys. Rev.*, **56,** 941.
Calderwood, J. H., and Cooper, R., 1953, *Proc. phys. Soc. Lond.* B, **66,** 73.
Calderwood, J. H., Cooper, R., and Wallace, A. A., 1953, *Proc. Inst. Radio Engrs, N.Y.*, **100,** 105.
Callen, H. B., 1949, *Phys. Rev.*, **76,** 1394.
Chuenkov, V., 1954, *Uspekhi Fiz. Nauk.*, **54,** 185.
Cooper, R., and Wallace, A. A., 1953, *Proc. phys. Soc. Lond.* B, **66,** 1113.
Cooper, R., Grossart, D. T., and Wallace, A. A., 1957, *Proc. phys. Soc. Lond.*, B, **70,** 169.
Davisson, J. W., 1946, *Phys. Rev.*, **70,** 685.
Fowler, J. F., 1956, *Proc. Roy. Soc.* A, **236,** 464.
Franz, W., 1939, *Z. Phys.*, **113,** 607; 1952, *Ibid.*, **132,** 285; *Encycl. Phys.*, 1956, **17,** 155.

Fröhlich, H., 1936, *Elektronentheorie der Metalle*, (Berlin: Springer); 1937, *Proc. roy. Soc.* A, **160,** 230; 1939, *Ibid.*, **172,** 94; 1947, *Ibid.*, **188,** 521, 532; 1949, *Theory of Dielectrics* (Oxford: University Press); 1952, *E. R. A. Report* L/T 277; 1954, *Advanc. Phys.*, **3,** 325.
Fröhlich, H., and Mott, N. F., 1939, *Proc. roy. Soc.* A, **171,** 496.
Fröhlich, H., Pelzer, H., and Zienau, S., 1950, *Phil. Mag.*, **41,** 221.
Fröhlich, H., and Seitz, F., 1950, *Phys. Rev.*, **79,** 526.
Fröhlich, H., and Paranjape, B. V., 1956, *Proc. phys. Soc. Lond.* B, **69,** 21.
Heller, W. R., 1951, *Phys. Rev.*, **84,** 1130.
von Hippel, A., 1935, *Ergebn. exakt. Naturw.*, **14,** 79; 1936, *Z. Phys.*, **98,** 580.
von Hippel, A., and Lee, G. M., 1941, *Phys. Rev.*, **59,** 824.
von Hippel, A., and Maurer, R. J., 1941, *Phys. Rev.*, **59,** 820.
Huang, K., and Rhys, A., 1950, *Proc. Roy. Soc.* A, **204,** 1950.
Kawamura, H., and Onuki, M., 1953, *J. phys. Soc. Japan*, **8,** 731.
Kawamura, H., Ohkura, H., and Kikuchi, T., 1954, *J. phys. Soc. Japan*, **9,** 541.
Keller, K. J., 1948, *Physica*, **14,** 475; 1951, *Ibid.*, **17,** 511.
Konorova, E. A., 1957, *J. exp. theor. Phys.*, **32,** 603.
Konorova, E. A., and Sorokina, L. A., 1957, *J. exp. theor. Phys.*, **32,** 143.
Lampert, M. A., 1956, *Phys. Rev.*, **103,** 1648.
Oakes, W. G., 1948, *J. Instn. elect. Engrs*, **95,** I, 36.
O'Dwyer, J. J., 1954, *Aust. J. Phys.*, **7,** 36, 405; 1957, *Proc. phys. Soc. Lond.* B, **70,** 761.
Paranjape, B. V., 1953, *E. R. A. Report*, L T 285; 1954, *Ibid.*, L T 307.
Pines, D., 1953, *Phys. Rev.*, **92,** 626.
Redfield, A., 1953, *Phys. Rev.*, **91,** 753.
Rose, A., 1955, *Phys. Rev.*, **97,** 1538.
Ryu, I., and Kawamura, H., 1954, *J. phys. Soc. Japan*, **9,** 438.
Seitz, F., 1948, *Phys. Rev.*, **73,** 549; 1949, *Ibid.*, **76,** 1376.
Simpson, J. H., 1949, *Proc. roy. Soc.* A, **197,** 269.
Stratton, R., 1957, *Proc. roy. Soc.* A, **242,** 355.
Veelken, R., 1955, *Z. Phys.*, **142,** 476, 544.
Vermeer, J., 1954, *Physica*, **20,** 313; 1956, *Ibid.*, **22,** 1247, 1257, 1269.
Vorob'ev, A. A., 1956, *Soviet Phys. Tech. Phys.*, **26,** 330.
Whitehead, S., 1950, *Dielectric Breakdown of Solids* (Oxford: University Press).
Zener, C., 1934, *Proc. roy. Soc.* A, **145,** 523.

Fig. 2

From *Arch. Biochem. Biophys.*

Southern beam mosaic virus protein crystal (molecular size, 245 Å). The large (110) face is in contact with a big (111) face at the upper left and a smaller (111) opposite. A third (111) at the upper right is separated from the (110) by a face of the {120} form. Preshadowed with Pd, backed by carbon. Magnification ×48 750. Labaw and Wyckoff (1955).

Fig. 3

From *Exp. Cell Res. Suppl.*

Plant protein crystal (molecular size, 140 Å). Preshadowed with Pd, backed by carbon. Magnification × 47 500. Labaw and Wyckoff (1955).

Fig. 4

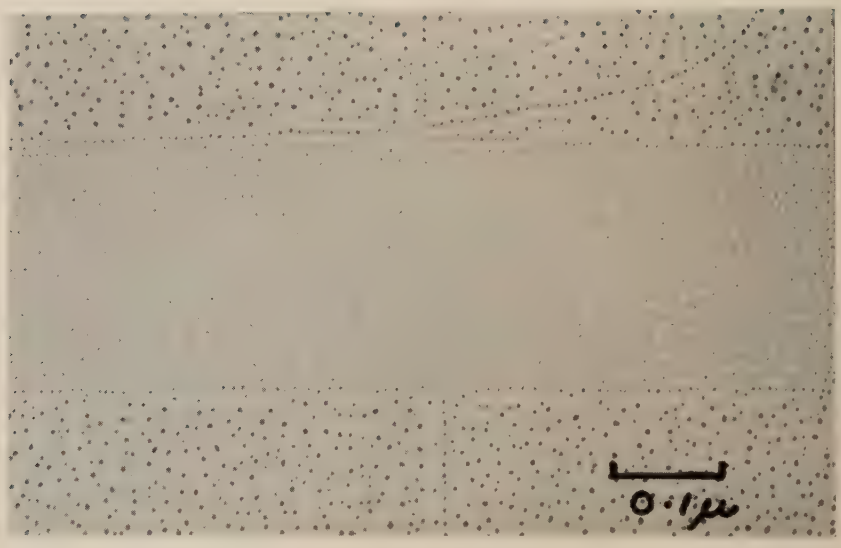
From *J. chem. Soc.*

Potash felspar crystal with surface steps revealed by shadowing with Au/Pd at $\tan^{-1}\frac{1}{4}$. Step height probably less than 25 Å. Magnification × 30 000. Gard *et al.* (1955).

Fig. 5

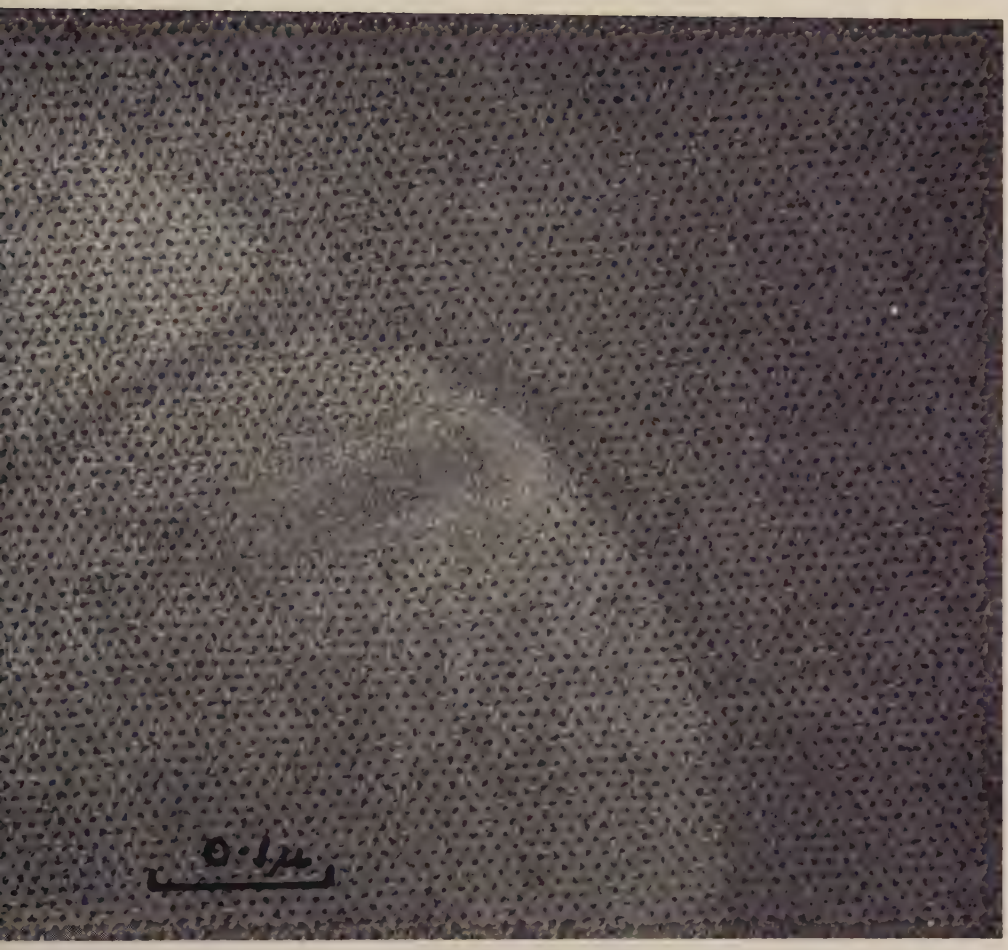
Cleavage and slip steps on rocksalt decorated with gold nuclei by vacuum evaporation of gold. Magnification × 40 000. Bassett (1958).

Fig. 6



Cleavage terrace on rocksalt crossed by slip step decorated with gold nuclei. 45 ∓ 2 steps in cleavage terrace. Terrace height 125 ∓ 6 Å. Many of the cleavage steps are 2·8 Å high. Magnification $\times 140\,000$. Bassett (1958).

Fig. 7



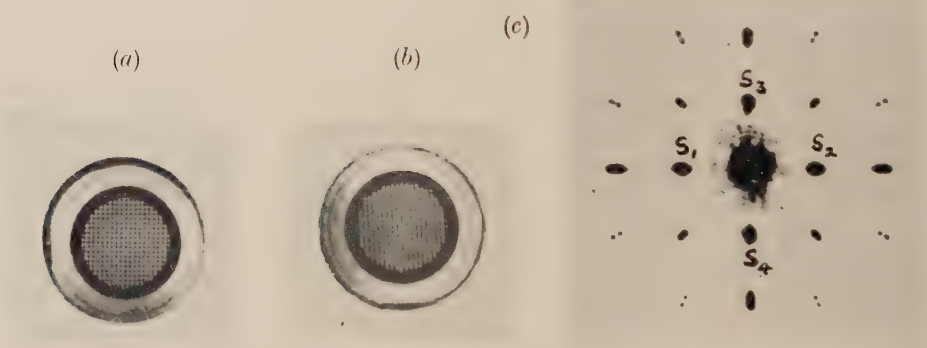
Microtome section of polyhedral inclusion body of insect virus showing macromolecular paracrystalline lattice. Virus occluded in the crystal does not disturb perfection of lattice. Darker areas in micrograph due to supporting film of nitrocellulose. Magnification ×233 000. Bergold and Fernandez–Moran (1957).

Fig. 10

From *Proc. roy. Soc. A.*

Portion of platinum phthalocyanine crystal showing perfect structure of (201) planes. Magnification $\times 1\,000\,000$. Menter (1956 a).

Fig. 11



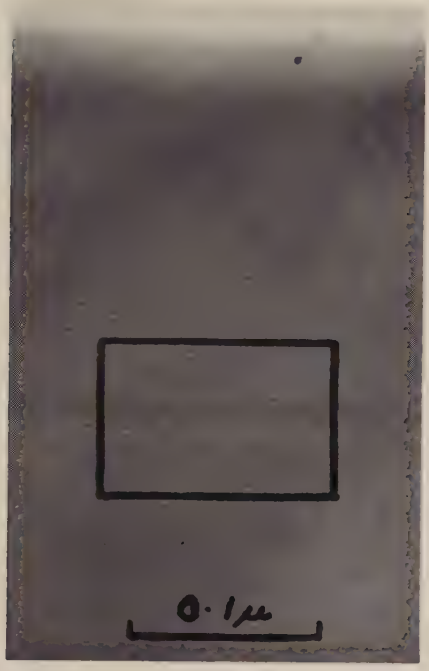
(*a*) Image of 200 t.p.i. grid with full aperture of lens. (*b*) Image of same grid with physical stop in back focal plane of lens to permit only beams S_0, S_1 and S_2 (fig. 11(*c*)) to contribute to image. (*c*) Diffraction pattern from grid recorded in back focal plane of lens.

Fig. 12

Diffraction pattern from Pt phthalocyanine crystal.

Fig. 13

(*a*)

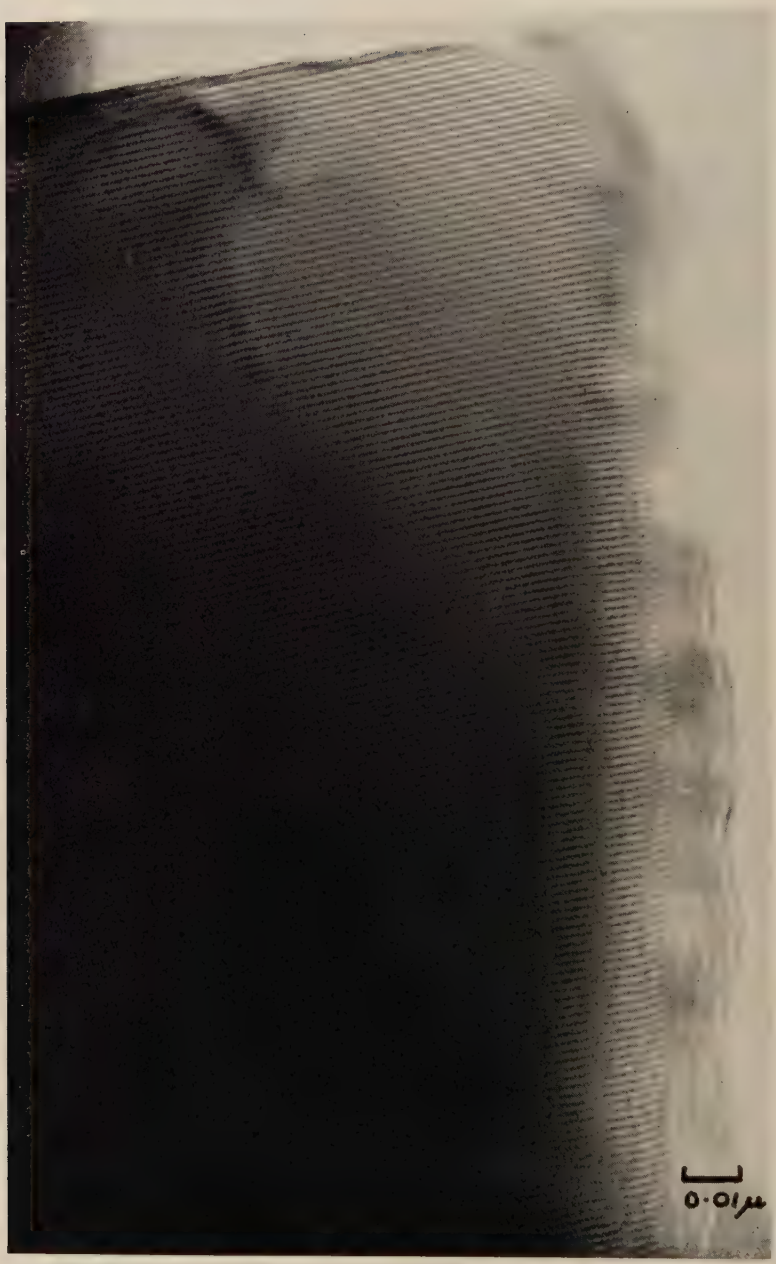


(*b*)

(*c*)

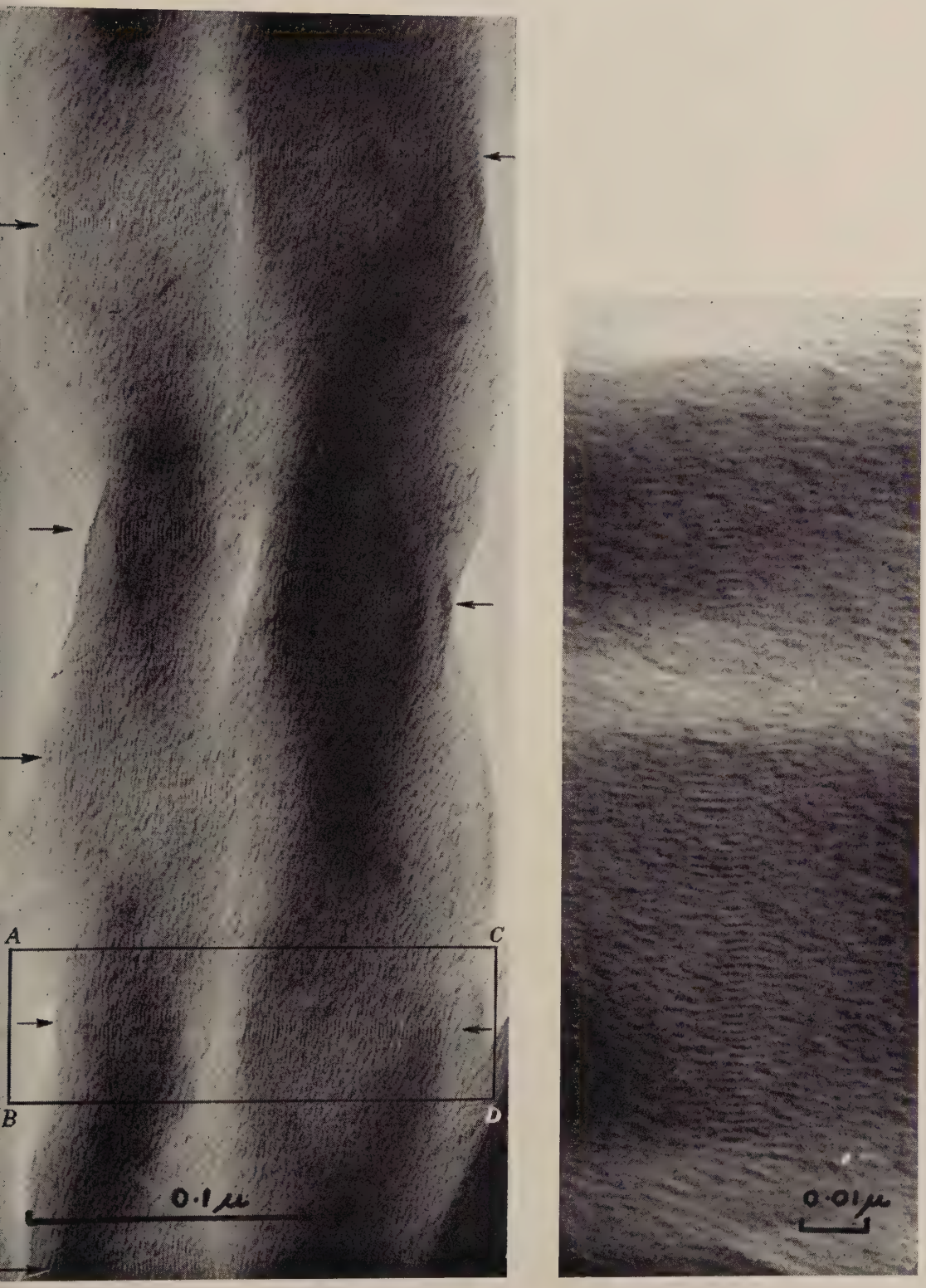
(*a*) Extinction contour in bent crystal of platinum phthalocyanine imaged with 10 μ objective aperture to exclude all diffracted beams from image except zero order. (*b*) Same crystal imaged with 30 μ aperture. Lattice is resolved (see enlargement (*c*) of enclosed region) only in those regions of crystal previously shown to be in reflecting orientation by presence of extinction contour. Magnification $\times 250\,000$. Menter (unpublished).

Fig. 14



Sodium faujasite crystal with (111) spacing resolved, $d=14{\cdot}4$ Å. Magnification $\times 750\,000$. Menter (1957).

Fig. 15

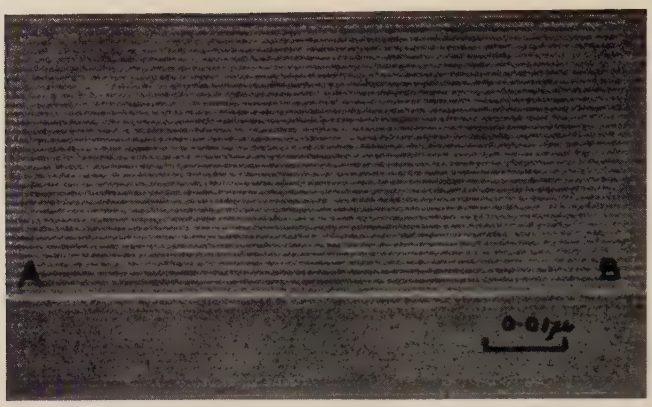


(*a*) (*b*)

From *Proc. roy. Soc. A.*

(*a*) Twisted crystals of copper phthalocyanine showing lattice resolved at regions indicated by arrows at intervals corresponding to twist of π in the crystal. Magnification $\times 500\,000$. (*b*) Enlargement of marked region in 16 (*a*). Magnification $\times 1\,000\,000$. Menter (1956 a).

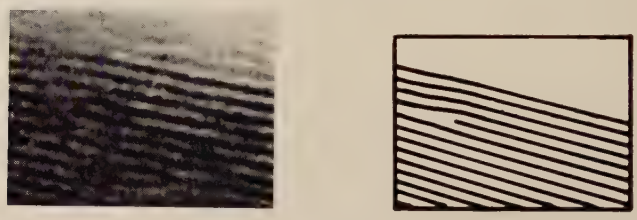
Fig. 16



From *Nature, Lond.*

Twinned copper phthalocyanine crystal. Lattice spacing changes abruptly from 12·5 Å to 9·7 Å across line AB. Magnification ×1 000 000. Komoda *et al.* (1958).

Fig. 18



From *Proc. roy. Soc. A.*

(*a*) (*b*)

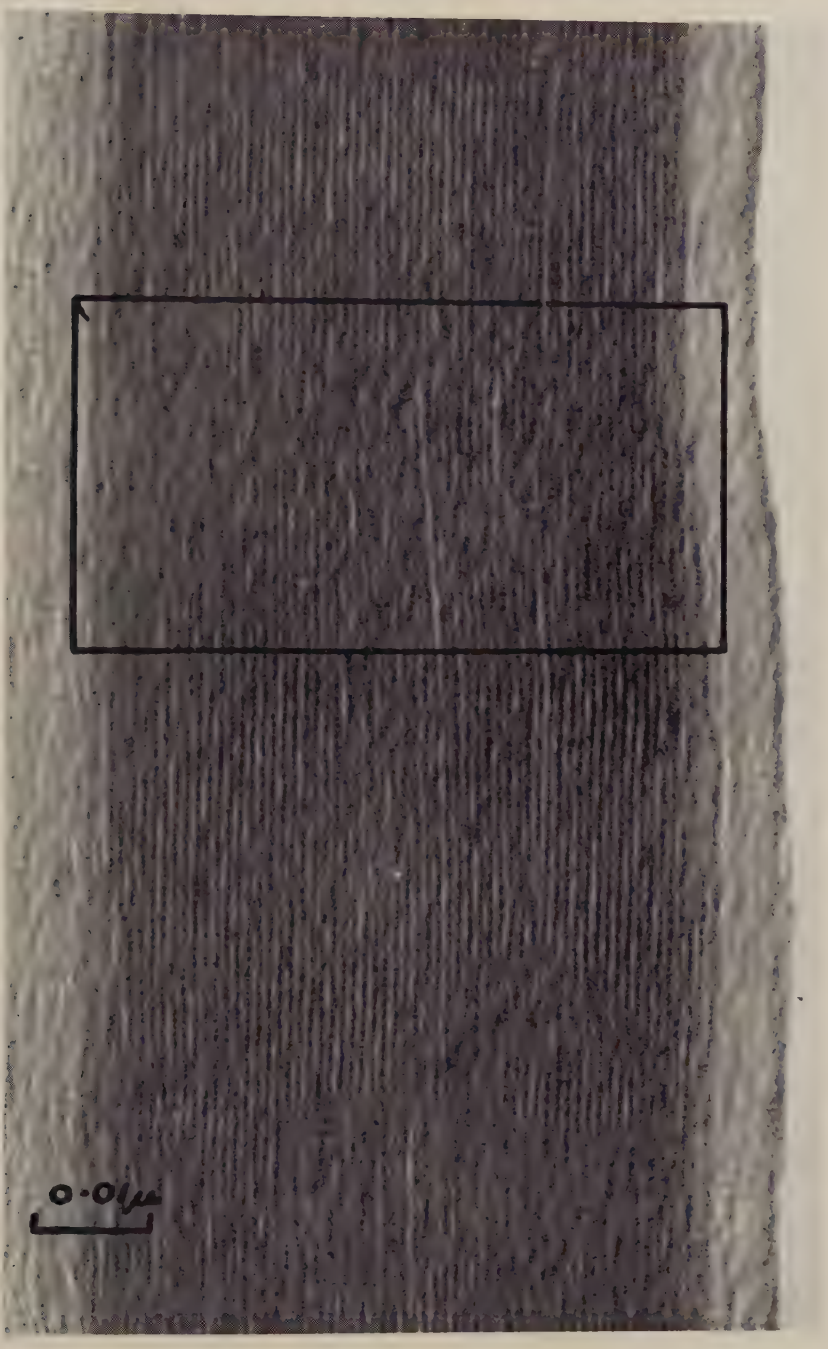
(*a*) Extra half line in resolved lattice of platinum phthalocyanine crystal. Magnification ×1 500 000. Menter (1956 a). (*b*) Sketch copied from micrograph in 19 (*a*).

Fig. 19

From *Times sci. Rev.*

Extra half line in resolved lattice of platinum phthalocyanine. The extra half line is here not unambiguously defined. It can be detected only by making a Burgers circuit around the diffuse area in the centre of the micrograph. Magnification ×1 000 000. Photo by Bassett (Menter 1956 b).

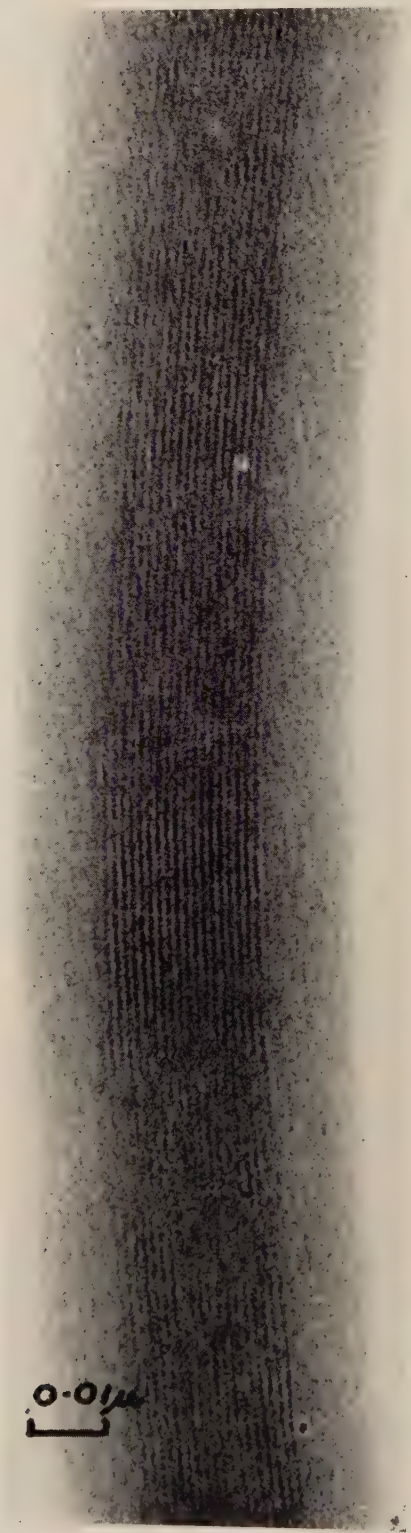
Fig. 21



From *Proc. roy. Soc. A.*

Copper phthalocyanine crystal with complex disturbance of lattice perfection in region marked. Magnification ×1 500 000. Menter (1956 a).

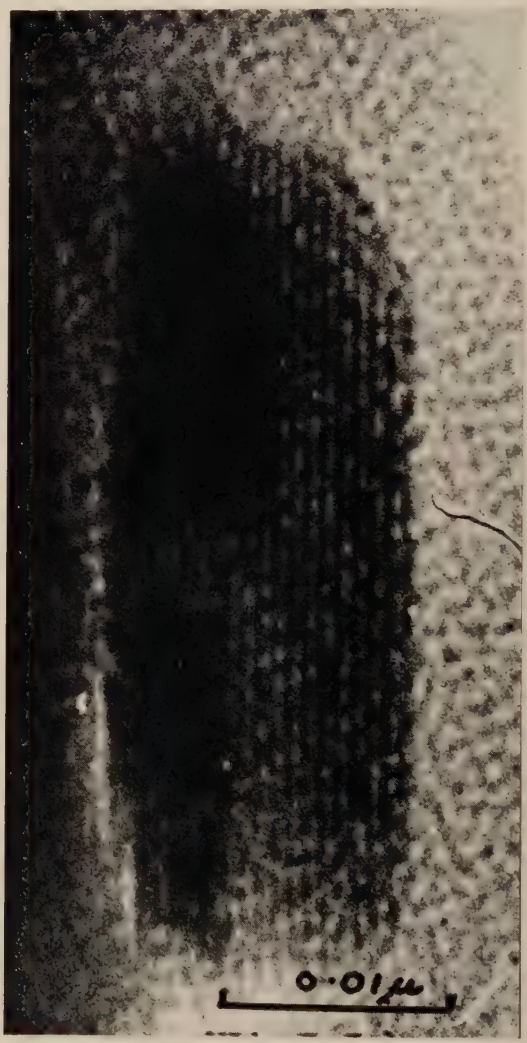
Fig. 22



From *Proc. roy. Soc. A.*

Strained crystal of platinum phthalocyanine which may be polygonal on account of edge dislocations with Burgers vector parallel to resolved planes. Magnification ×1 000 000. Menter (1956 a).

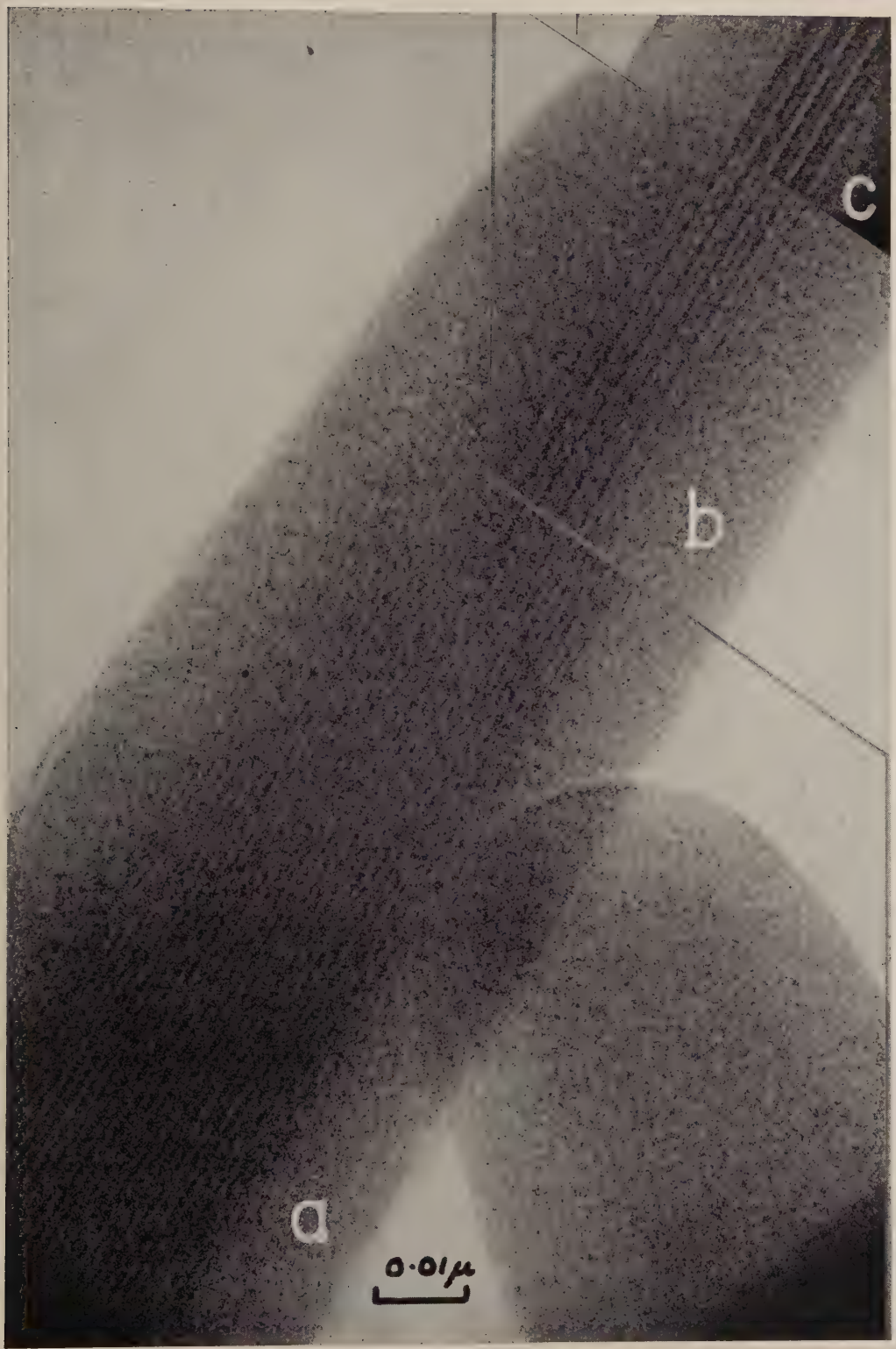
Fig. 25



From *Phil. Mag.*

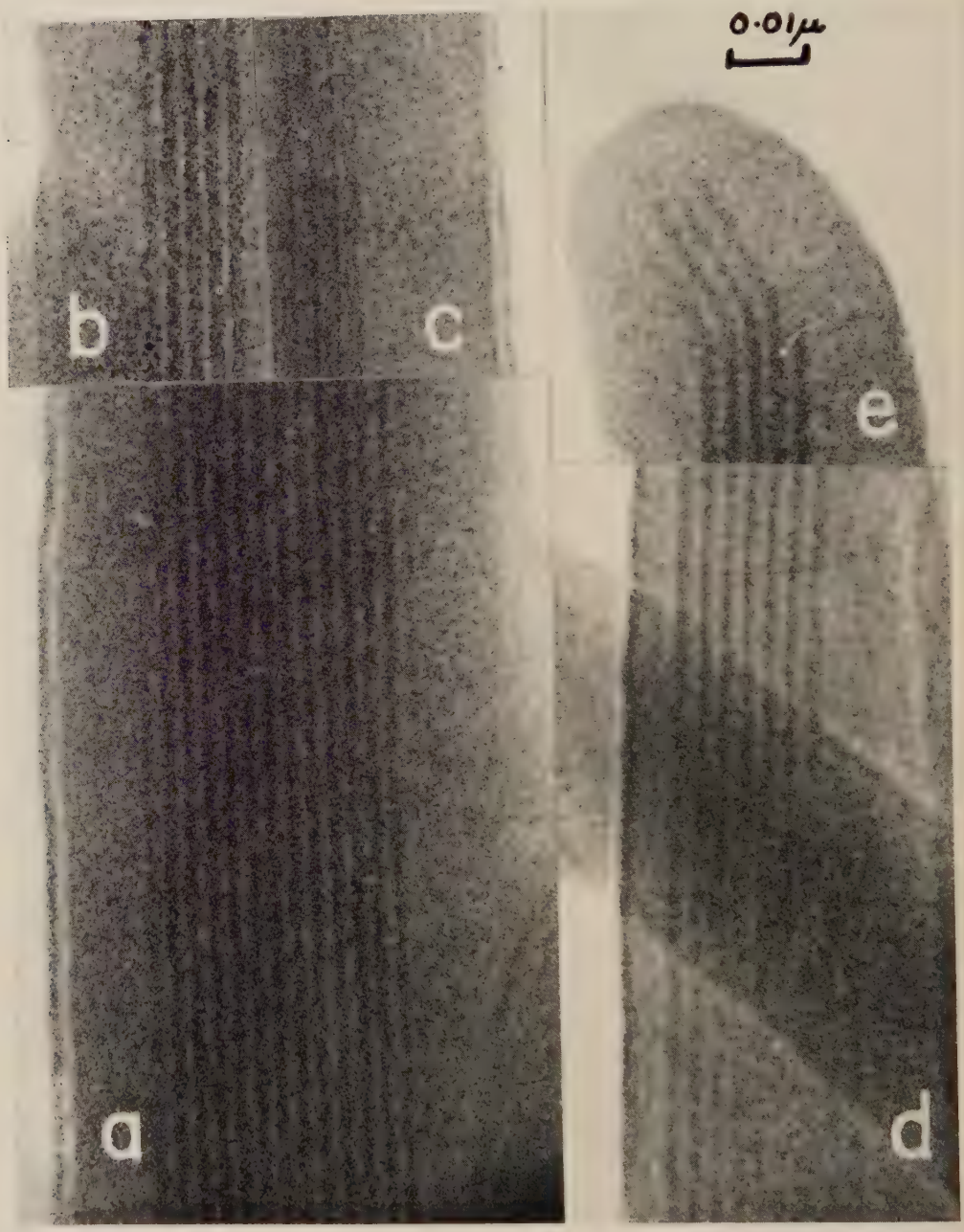
Molybdenum trioxide crystal. (020) spacing of 6·9 Å resolved. This is the smallest lattice spacing so far resolved directly. Magnification ×3 000 000. Bassett and Menter (1957).

Fig. 23



Indanthrene scarlet R crystals. Resolved spacing 15·4 Å (*a*), also showing intensity modulation (*b*), (*c*). Magnification ×1 360 000. Labaw and Wyckoff (1957).

Fig. 24



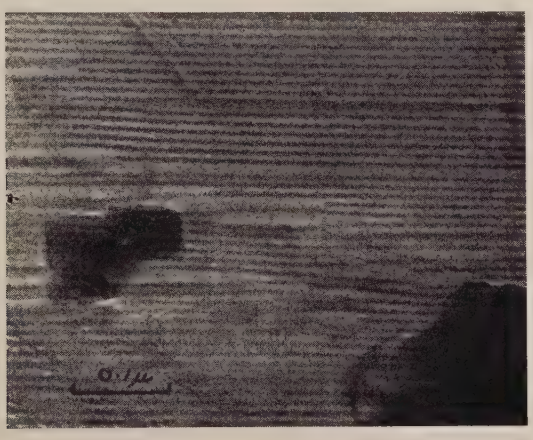
From *Proc. nat. Acad. Sci. Wash.*

Indanthrene scarlet R crystals showing 28·1 Å spacing (*a*), and splitting into 14·0 Å spacing (*b*), (*c*). (*d*) is another crystal with the 28·1 Å spacing resolved and (*e*) shows the end of this crystal where the lattice direction changes abruptly, possibly due to twinning. Magnification ×1 000 000. Labaw and Wyckoff (1957).

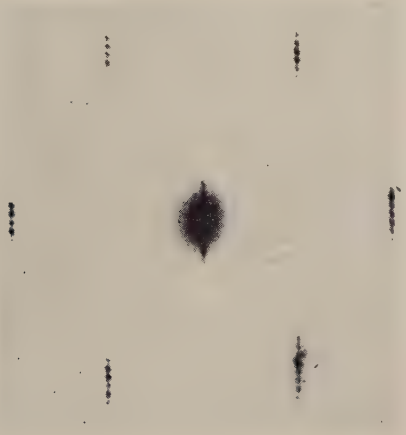
Fig. 26

Faujasite crystal viewed along ‹110› with two sets of (111) spacings resolved (14·4 Å). Magnification 750 000. Menter (1957).

Fig. 27



(*a*)

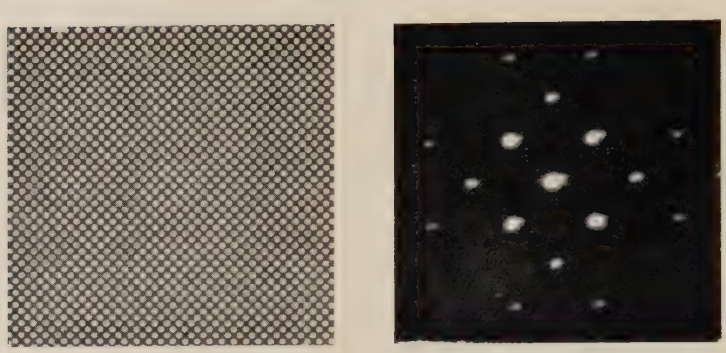


(*b*)

From *Acta cryst.*

(*a*) 100 Å superlattice spacing resolved in antigorite. Magnification ×125 000. (*b*) Part of diffraction pattern from antigorite crystal showing closely spaced satellites about main reflections. Similar satellites are thought to form about the zero order beam and recombine to form the image shown in fig. 30 (*a*). Brindley *et al.* (1958).

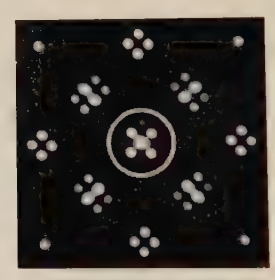
Fig. 32



(*a*) (*b*)

(*a*) Elementary grating used in analogue. (*b*) Diffraction pattern from grating shown in fig. 32 (*a*).

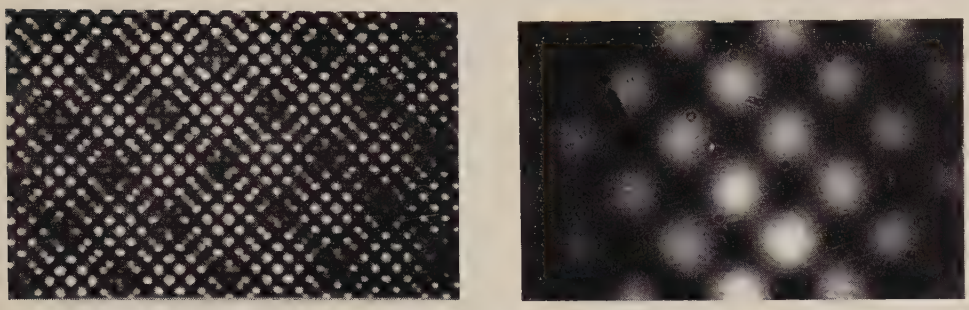
Fig. 33



From *Proc. roy. Soc.*

Double diffraction pattern from two superposed gratings of type shown in fig. 32 (*a*). Spacings 220/inch and 100/inch. Pattern consists of two primary patterns with square symmetry and doubly diffracted beams around the primaries.

Fig. 34

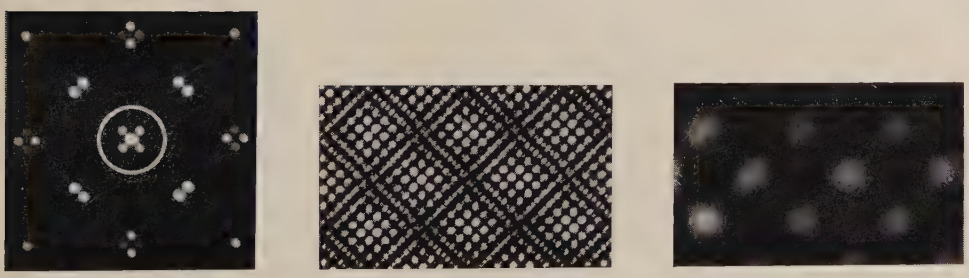


From *Proc. roy. Soc.*

(*a*) (*b*)

(*a*) Image of superposed gratings with full lens aperture. (*b*) Same with stop in back focal plane of lens to permit only doubly diffracted beams around zero order (those within circle in fig. 33) to pass through to image plane.

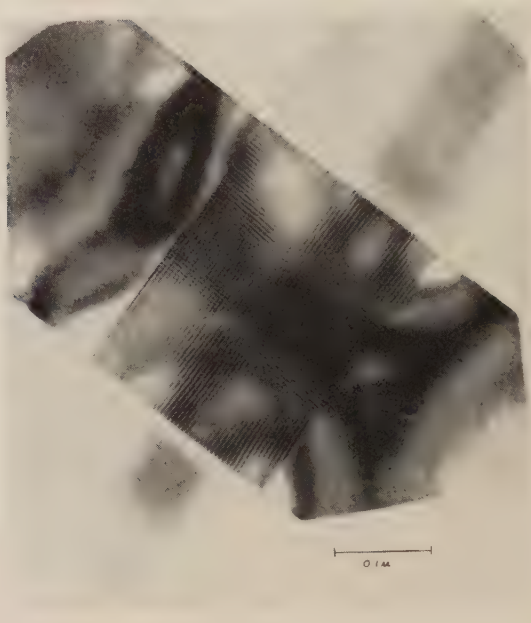
Fig. 35



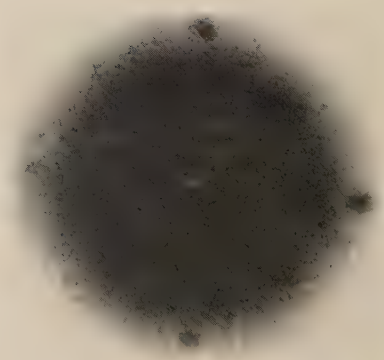
(*a*) (*b*) (*c*)

(*a*) Double diffraction pattern from two identical superposed gratings with relative rotation of 10°. (*b*) Image of superposed gratings with full lens aperture. (*c*) Same with lens stopped down as in fig. 34 (*b*).

Fig. 36



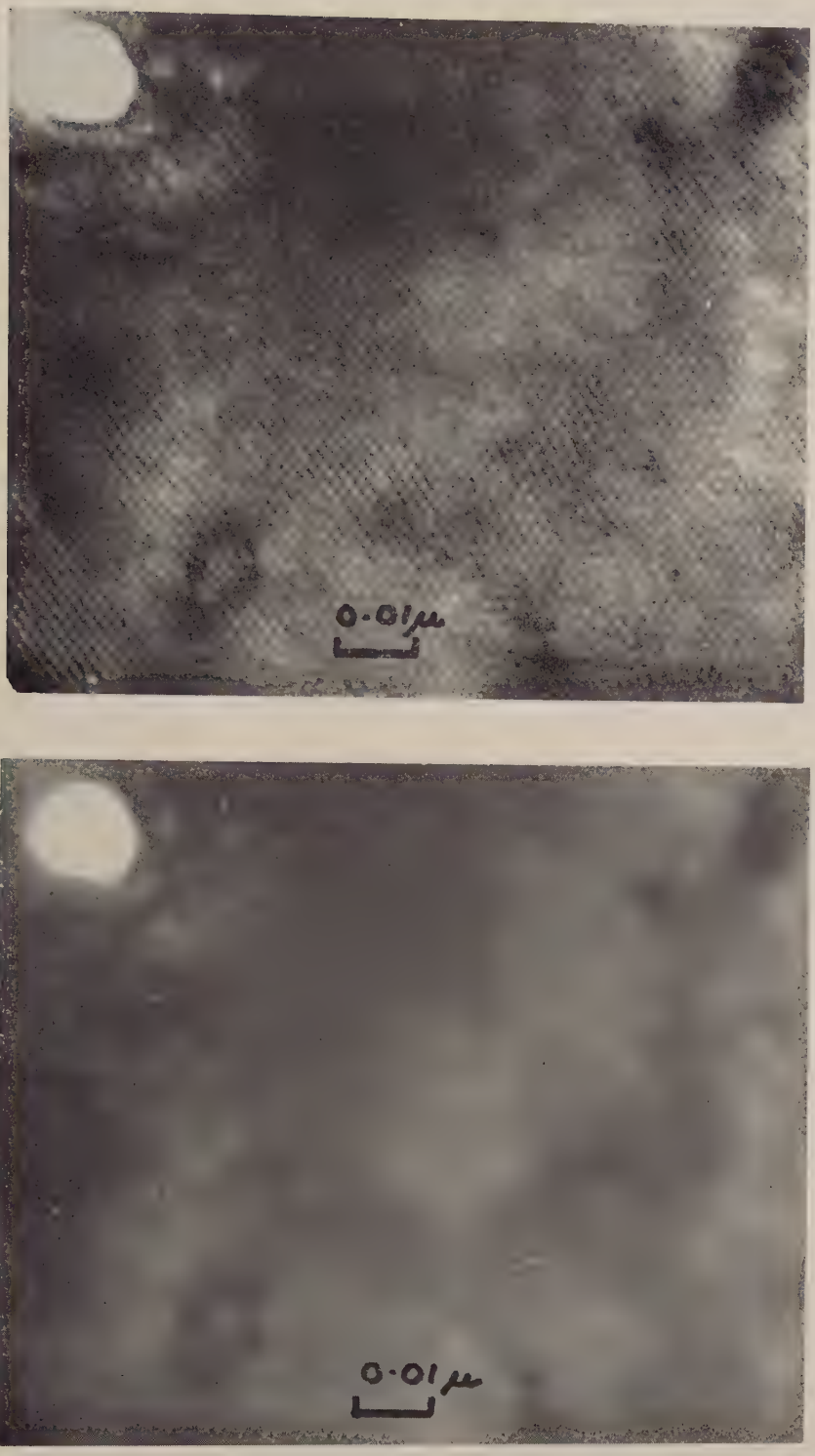
(*a*)



(*b*)

(*a*) Crossed crystals of molybdenum oxide with *a* almost parallel to *b*, showing moiré pattern in region of overlap. Magnification ×125 000. Dowell *et al.* (1956). (*b*) Doubly diffracted beams close to zero order beam from overlapping molybdenum oxide crystals similar to those shown in fig. 40. Dowell *et al.* (1957).

Fig. 37

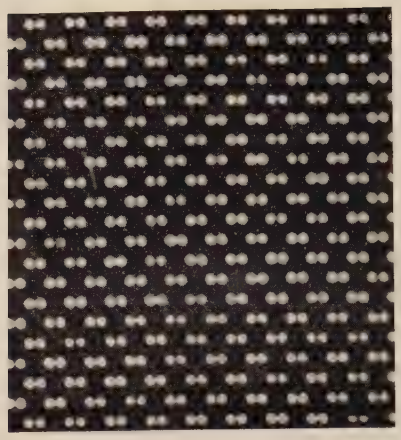


(*a*)

(*b*)

(*a*) Moiré pattern from parallel single crystals of copper and gold imaged with 50 μ objective aperture. (*b*) Same field imaged with 5 μ aperture to exclude all diffracted beams except zero order from image. The moiré pattern is not resolved, only non-periodic detail is seen. Magnification $\times 1\,000\,000$. Photo by Bassett (unpublished).

Fig. 38

(*a*)   (*b*)

Elementary grating (*a*) used to demonstrate production of Patterson distribution (*b*) by superposition on an identical grating with a small rotation about an axis normal to the page. Dowell *et al.* (1957).

Fig. 40

(*a*)   (*b*)

From *Nature, Lond.*

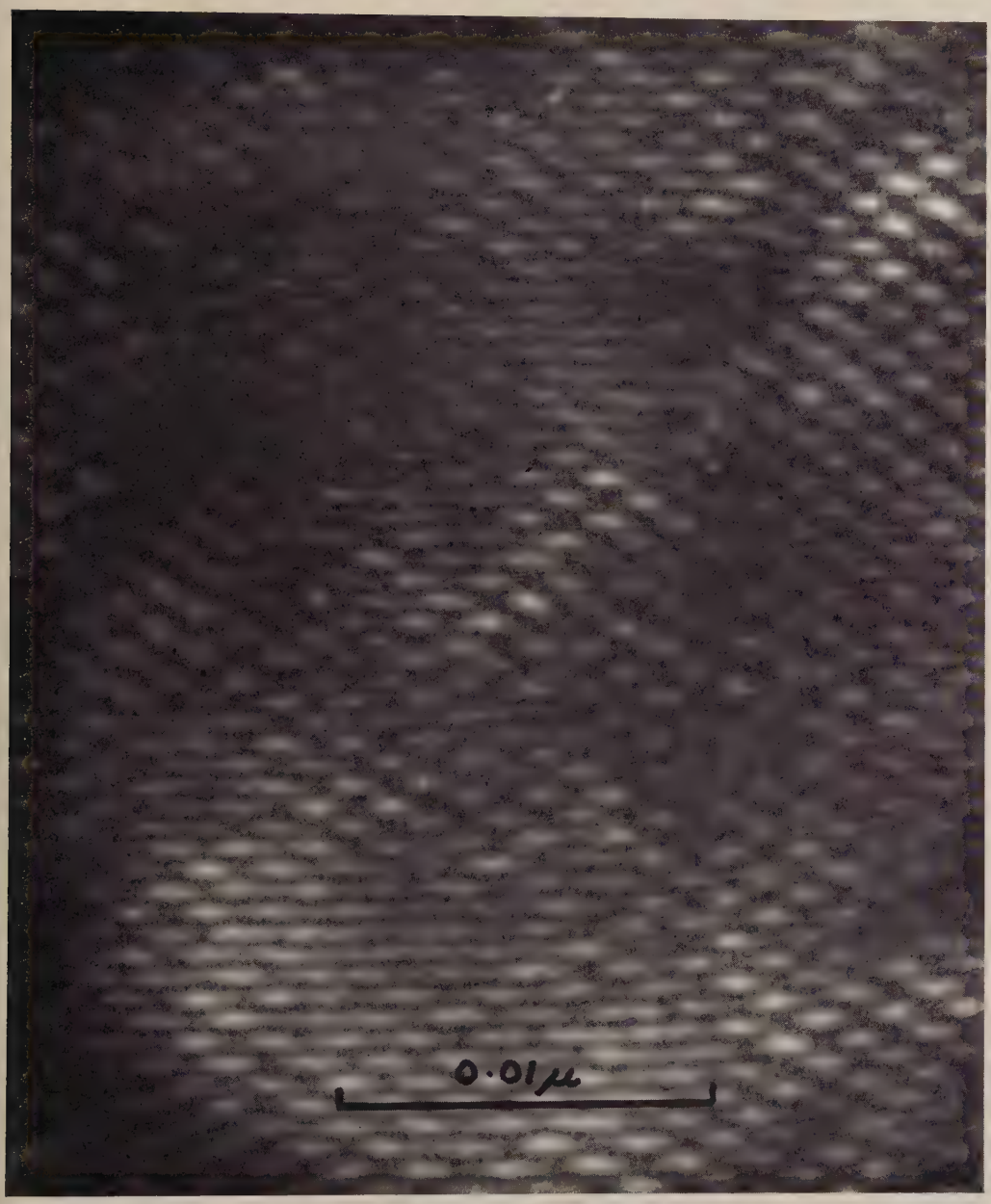
(*a*) Diffraction pattern from gold single crystal film grown by vacuum evaporation. ⟨111⟩ normal to plane of page. Pashley (1958). (*b*) Double diffraction pattern from specimen consisting of single crystal of nickel grown epitaxially (in parallel orientation) on gold single crystal. Pashley *et al.* (1957).

Fig. 45

From *Proc. roy. Soc.*

Double diffraction pattern from two gold films separately prepared and superposed with small relative rotation. Bassett *et al.* (1958).

Fig. 39



From *Proc. roy. Soc.*

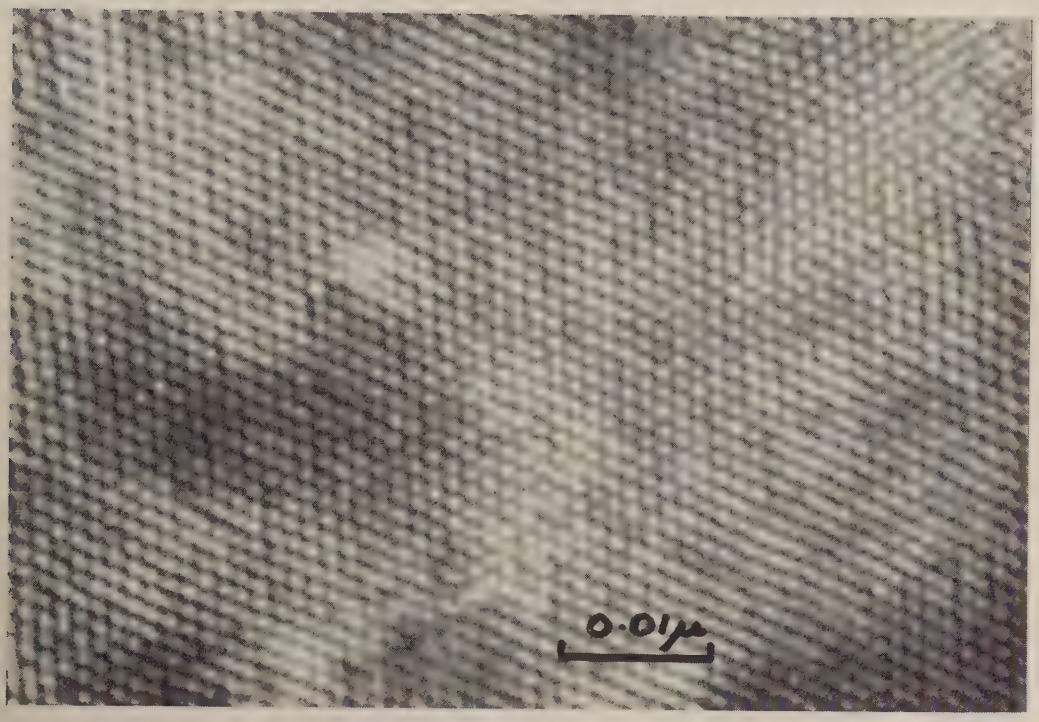
{422} type moiré pattern from Ni/Au single crystal specimen. Measured spacing 5·8 Å, theoretical spacing 5·3 Å. This is the smallest periodic structure so far resolved in the electron microscope. Magnification ×5 000 000. Bassett *et al.* (1958).

Fig. 41

From *Proc. roy. Soc.*

Moiré pattern from specimen consisting of palladium single crystal grown epitaxially upon a gold single crystal. {422} and {220} patterns are seen. The pattern varies from one region to another due to variations in reflecting conditions arising from buckling of specimen. Magnification $\times 400\,000$. Bassett *et al.* (1958).

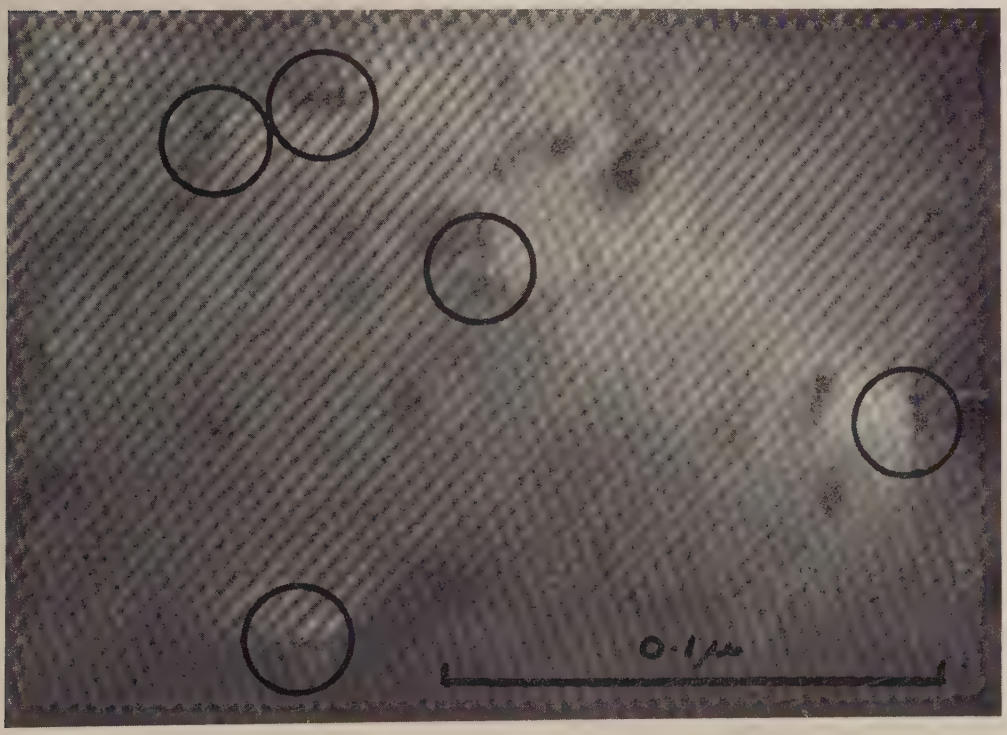
Fig. 42



From *Proc. roy. Soc.*

Moiré pattern from Cu/Au parallel single crystals showing large area of dot pattern. At least two and possibly three (220) reflections are contributing to this image. Magnification ×2 000 000. Bassett *et al.* (1958).

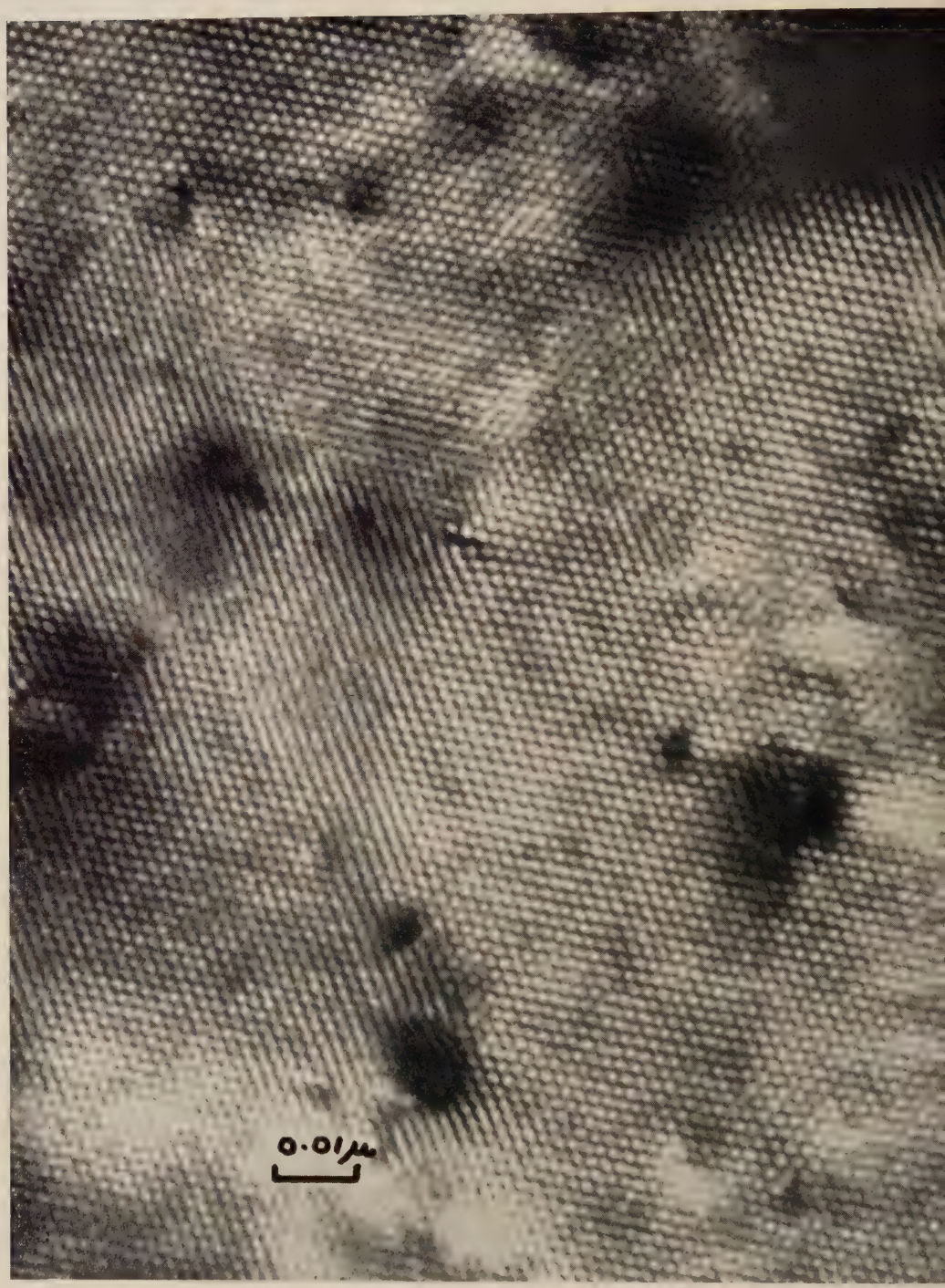
Fig. 43



From *Proc. roy. Soc.*

Moiré pattern from two gold films separately prepared and superposed with a relative rotation of ~2°. Magnification ×640 000. Bassett *et al.* (1958).

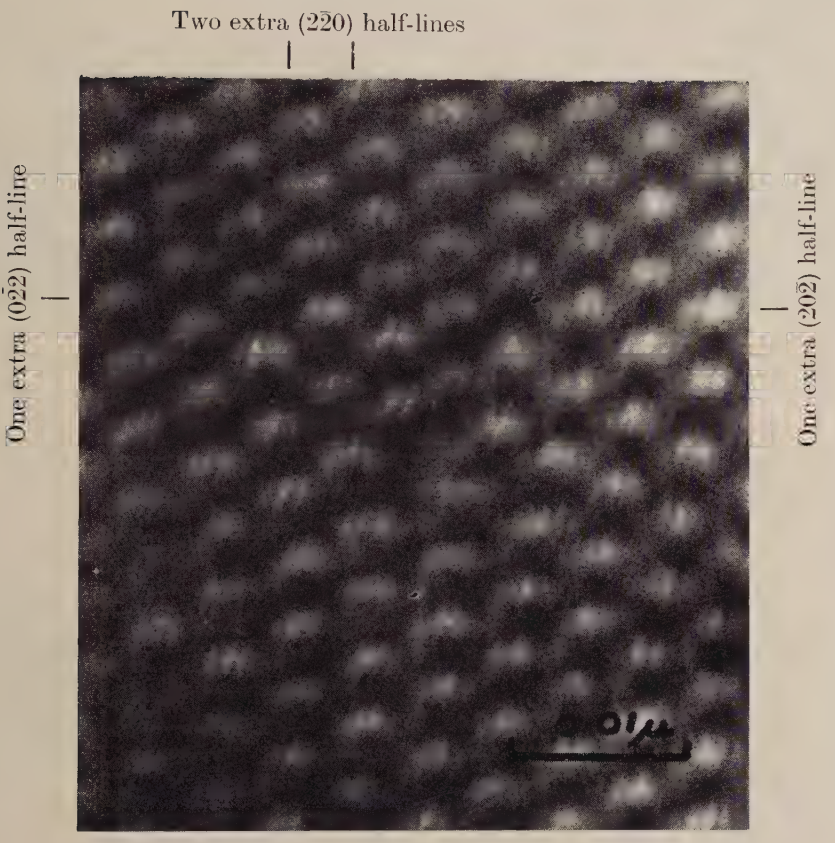
Fig. 44



From *Proc. roy. Soc.*

Moiré pattern from separately prepared and superposed single crystal films of gold and palladium. Magnification ×1 200 000. Bassett *et al.* (1958).

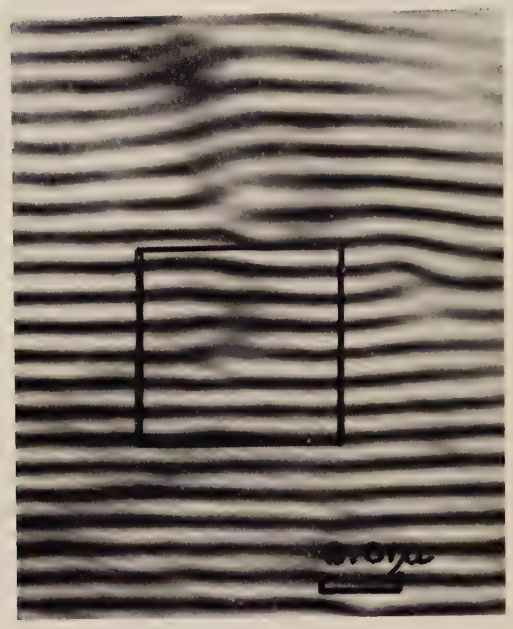
Fig. 47



From *Proc. roy. Soc.*

Dislocation in moiré pattern from Au/Pd parallel single crystal specimen. The Burgers vector lies in the plane of the specimen in a horizontal direction. Magnification ×2 300 000. Bassett *et al.* (1958).

Fig. 48

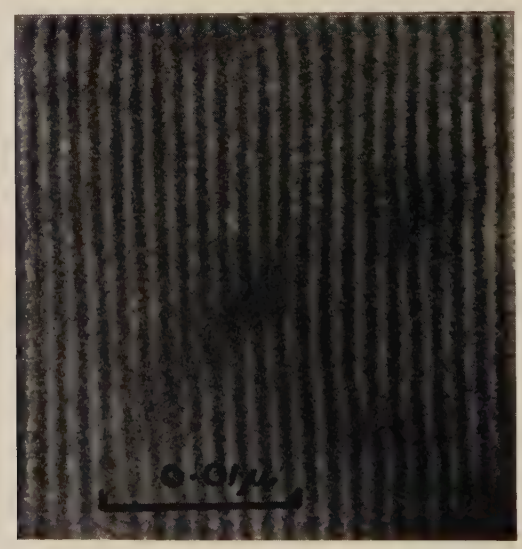


From *Proc. roy. Soc*

(*a*) (*b*)

(*a*) Moiré pattern from Au Pd parallel single crystals. No extra half lines in enclosed region. (*b*) Same 15 seconds later. It is supposed that in (*a*) there was a dislocation line running from top to bottom of the gold specimen which was copied by the Pd during growth of the latter on the Au. As long as these two dislocations of the same sign remain immediately above each other they are not seen in the moiré pattern. In (*b*) the two lines must have become displaced sideways and appear as two extra half lines of opposite sign in the moiré pattern. Magnification $\times 1\,000\,000$. Bassett *et al.* (1958).

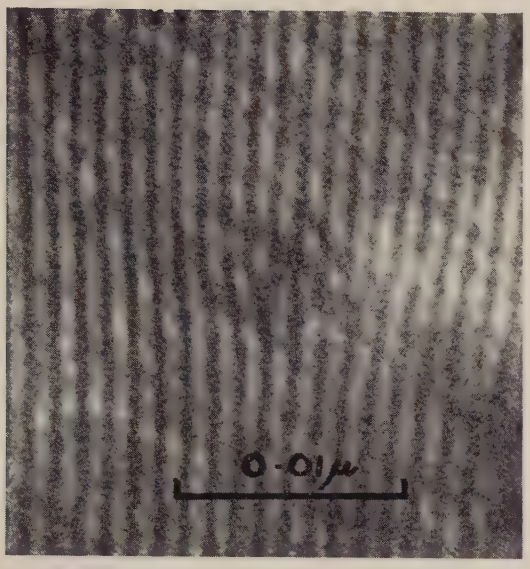
Fig. 49



From *Proc. roy Soc*

Stacking fault in moiré pattern from separately prepared superposed single crystal films of gold and palladium. A displacement of $+\frac{1}{3}$ spacing is observed at one end of fault and $+\frac{2}{3}$ at other end. Magnification $\times 2\,600\,000$. Bassett *et al.* (1958).

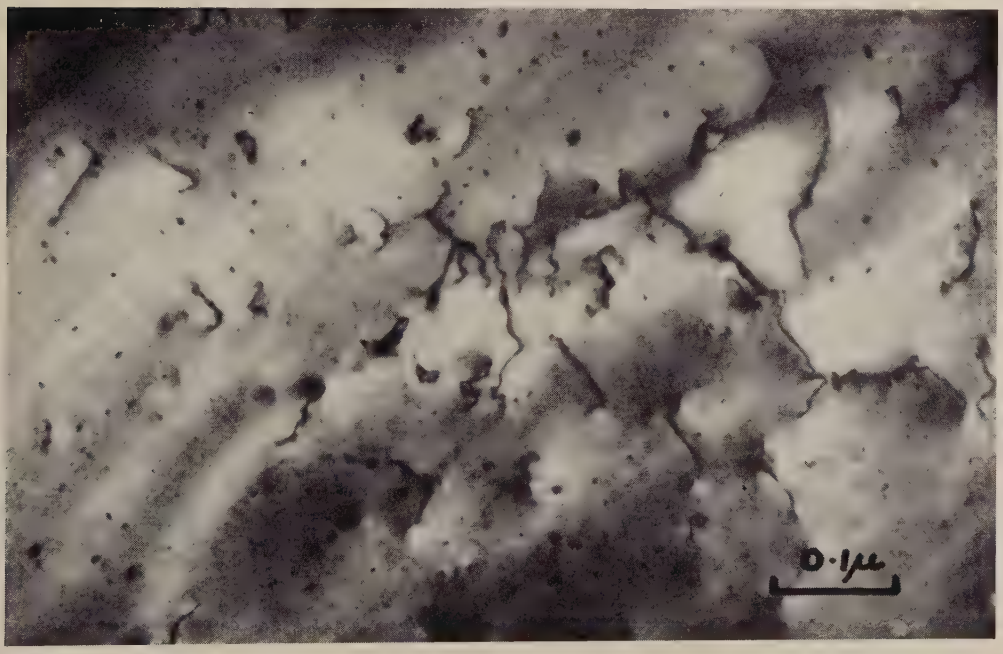
Fig. 50



From *Proc. roy. Soc.*

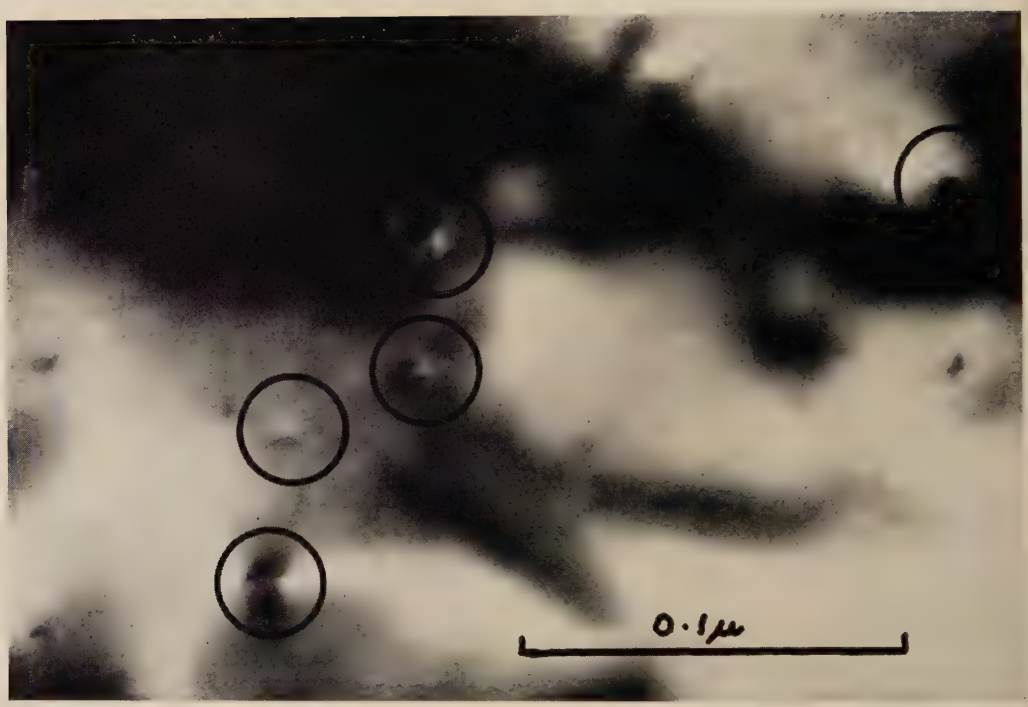
Two extra half lines in moiré pattern from parallel single crystals of copper and gold arising from extended dislocation in one of the crystals. Separation of partial dislocations ~40 Å. Magnification ×3 000 000. Bassett *et al.* (1958).

Fig. 51



Dislocation lines in thin section of platinum cut on ultramicrotome, revealed by diffraction contrast after Hirsch *et al.* (1956). Magnification ×160 000. Menter (unpublished).

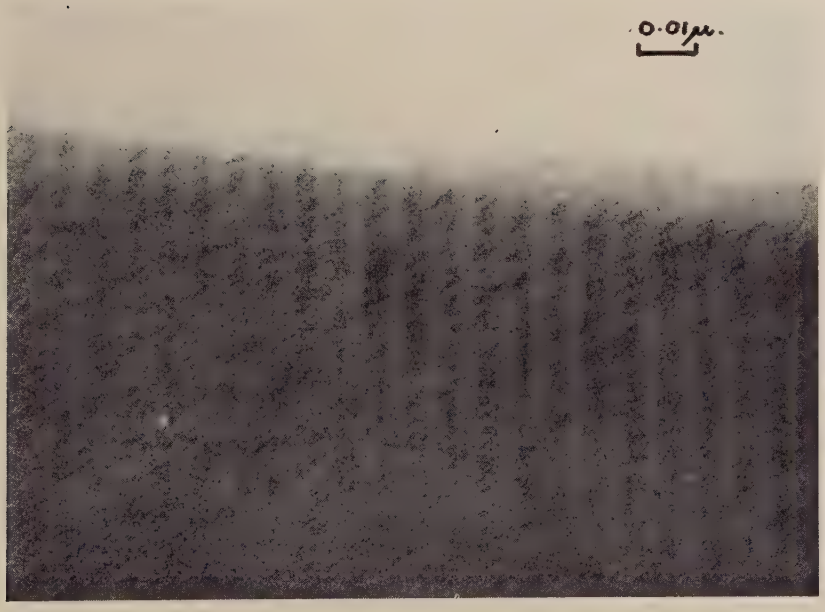
Fig. 52



Single crystal films of gold, ~300 Å in thickness. Slip planes are inclined at 70° to film surface. Dislocation lines are seen almost end on and appear as characteristic black dots. Magnification ×500 000. Pashley (1958).

Fig. 53

(*a*)



(*b*)

(*a*) Overlapping crystals of platinum phthalocyanine showing (20$\bar{1}$) spacing resolved in both crystals and moiré pattern formed in region of overlap. 50 μ objective aperture. Magnification $\times 750\,000$. (*b*) Same area with 10 μ objective aperture to exclude diffracted beams from (20$\bar{1}$) planes. Moiré pattern only is resolved. Menter (unpublished).